

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 20:22:25 ; Search time 104 Seconds
(without alignments)
1347.332 Million cell updates/sec

Title: US-09-759-207-2
Perfect score: 2842
Sequence: 1 MLRSKALPPPLMLLLGP.....LPFYSFVIRAKVACT 543

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL.23.1
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriophage:
17: sp_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2838	99.9	543	4 Q9Y251	Q9Y251 homo sapien
2	2817	99.1	545	4 Q9UD39	Q9UD39 homo sapien
3	2282	80.3	545	6 Q9MY10	Q9MY10 bos taurus
4	2150	75.7	535	11 Q8K3K3	Q8K3K3 mus musculus
5	2123	74.7	536	11 Q9QZFB	Q9QZFB ratuus norv
6	1645.5	57.9	523	13 Q9QYK5	Q9QYK5 gallus gall
7	1154.5	40.6	592	4 Q9HB37	Q9HB37 homo sapien
8	1146.5	40.3	592	4 Q8WQ02	Q8WQ02 homo sapien
9	1015.5	35.7	548	4 Q8WQ01	Q8WQ01 homo sapien
10	936.5	33.0	534	4 Q9HB38	Q9HB38 homo sapien
11	897.5	31.6	480	4 Q9HB39	Q9HB39 homo sapien
12	696	24.5	515	5 Q8T108	Q8T108 bombyx mori
13	416	14.6	521	10 Q9SDA1	Q9SDA1 arabidopsis
14	392.5	13.8	543	10 Q9FF10	Q9FF10 arabidopsis
15	381	13.4	527	10 Q9LRC8	Q9LRC8 oryza sativ
16					Q9LRC8 scutellaria

17	363	12.8	536	10 Q9FZP1	Q9FZP1 arabidopsis
18	352.5	12.4	536	10 Q9FLX8	Q9FLX8 arabidopsis
19	352.5	12.4	539	10 Q8L608	Q8L608 arabidopsis
20	169.5	6.0	190	10 Q82604	Q82604 arabidopsis
21	160	5.6	935	5 Q9VE79	Q9VE79 drosophila
22	130.5	4.6	493	17 Q9HK01	Q9HK01 thermoplasma
23	122.5	4.3	408	3 Q9HE21	Q9HE21 phanerochaete
24	122.5	4.3	408	3 Q9HE22	Q9HE22 phanerochaete
25	116.5	4.1	617	12 Q40996	Q40996 measles vir
26	114	4.0	398	16 Q8F410	Q8F410 leprospira
27	113.5	4.0	617	12 Q83295	Q83295 measles vir
28	113.5	4.0	1829	2 Q9KH44	Q9KH44 pantoea agg
29	113	4.0	390	17 Q8RPH7	Q8RPH7 methanosarc
30	112.5	4.0	617	12 Q83647	Q83647 measles vir
31	111.5	3.9	493	17 Q979W0	Q979W0 thermoplasma
32	111.5	3.9	575	10 Q43855	Q43855 vicia faba
33	111.5	3.9	617	12 Q40991	Q40991 measles vir
34	111.5	3.9	617	12 Q96VH5	Q96VH5 measles vir
35	111	3.9	670	10 Q9M090	Q9M090 arabidopsis
36	111	3.9	2319	3 Q96U00	Q96U00 neurospora
37	110.5	3.9	475	5 Q8SU17	Q8SU17 encephalito
38	109.5	3.9	617	12 Q911P6	Q911P6 measles vir
39	109.5	3.9	617	12 Q910N9	Q910N9 measles vir
40	109.5	3.9	879	16 Q8XCP4	Q8XCP4 escherichia
41	109	3.8	617	12 Q96VH6	Q96VH6 measles vir
42	108.5	3.8	411	16 P72895	P72895 synchocyst
43	108.5	3.8	617	12 Q11381	Q11381 measles vir
44	107.5	3.8	500	16 Q9ASU0	Q9ASU0 cauliobacter
45	107.5	3.8	617	12 Q8Q593	Q8Q593 measles vir

ALIGNMENTS

RESULT 1
Q9Y251 PRELIMINARY: PRT: 543 AA.
AC Q9Y251
DT 01-NOV-1999 (TRENBLREL 12, Created)
DT 01-NOV-1999 (TRENBLREL 12, Last sequence update)
DT 01-OCT-2002 (TRENBLREL 22, Last annotation update)
DE HEPRANASE.
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis,"
RT Nat. Med. 5:803-809(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Vlodavsky I., Friedman Y., Elkin M., Aingorn H., Atzmon R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Perez T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and
RT metastasis,"
RT submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.,
RA "Human heparanase. Purification, characterization, cloning, and
RT expression,"
RT J. Biol. Chem. 274:24153-24160(1999).
RN [4]
RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;
 RA MEDLINE=9935379; PubMed=10405343;
 RX Kneissel P.H., Holmes J.D., Ludwig D., Patel S., Navarro E.C.,
 Seddon A.P., Giorgio N.A., Bohlen P.,
 RT "Cloning and Functional Expression of a Human Heparanase Gene."
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
 DR EMBL: AF144325; AAD45379.1;
 DR EMBL: AF144325; AAD41442.1;
 DR EMBL: AF155510; AAD54941.1;
 DR EMBL: AF152376; AAD45669.1;
 DR Genbank: HGNC:5164; HPSE.
 DR InterPro: IPR005199; Glyco_hydro_79n.
 DR Pfam: PF03662; Glyco_hydro_79n; I.
 DR SEQUENCE 543 AA; 61176 MW; AD262EC67334AB2 CRC64;
 SQ

Query Match 99.9%; Score 2838; DB 4; Length 543;
 Best Local Similarity 99.8%; Pred. No. 3,4e-218;
 Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLRSKRALPPPLMLLLGPGPLSPGALPRPAQADVDVDFPQEPHLVSPSFLSVT	60
DB	1	MLRSKRALPPPLMLLLGPGPLSPGALPRPAQADVDVDFPQEPHLVSPSFLSVT	60
QY	61	IDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYMS	120
DB	61	IDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYMS	120
QY	121	QVQVODICKGSIPEPDVEEKLRLWPYQOULLREHYQKKFKNSTYSRSSVDVLYTFANCS	180
DB	121	QVQVODICKGSIPEPDVEEKLRLWPYQOULLREHYQKKFKNSTYSRSSVDVLYTFANCS	180
QY	181	GDLIFGLNALRTADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFINCS	240
DB	181	GDLIFGLNALRTADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFINCS	240
QY	241	QGLGEDYIQLHKLKRSKTFKNAKLYGPDVQGPRAKTKAMLSFLKAGGEVYDSTVMHHYVL	300
DB	241	QGLGEDYIQLHKLKRSKTFKNAKLYGPDVQGPRAKTKAMLSFLKAGGEVYDSTVMHHYVL	300
QY	301	NGRTATREDPLNDVDLFISSVQKFFOVVESTRPCKTWLGETSSAYGGAAPLSDTFA	360
DB	301	NGRTATREDPLNDVDLFISSVQKFFOVVESTRPCKTWLGETSSAYGGAAPLSDTFA	360
QY	361	AGFMWLDKLGSLARMGIEVVMRQVFPAGAGNYHLVDENFDPPLPDYWLSTLLFKLVGKTVKM	420
DB	361	AGFMWLDKLGSLARMGIEVVMRQVFPAGAGNYHLVDENFDPPLPDYWLSTLLFKLVGKTVKM	420
QY	421	ASVQSGSKRRKRLRYLHCTNTDNPYKESGDLTYAINLHVTKYLRPLYPFSNKQVDKYL	480
DB	421	ASVQSGSKRRKRLRYLHCTNTDNPYKESGDLTYAINLHVTKYLRPLYPFSNKQVDKYL	480
QY	481	RPLGPHGLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPAFSSFFVIIRNAKVA	540
DB	481	RPLGPHGLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPAFSSFFVIIRNAKVA	540
QY	541	ACI 543	
DB	541	ACI 543	

RESULT 2
 Q9MUY39 PRELIMINARY; PRT; 545 AA.

AC Q9MUY39; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Heparanase.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN NCB1_TaxID=9606;
 RX TISSUE=Placenta;
 RC

RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=20229546; PubMed=10764835;
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.,
 RT "Heparanase expression in invasive trophoblasts and acute vascular
 damage."
 RL Glycobiology 10:467-475(2000).
 DR EMBL: AF084467; AAD54516.1;
 DR InterPro: IPR005199; Glyco_hydro_79n.
 DR Pfam: PF03662; Glyco_hydro_79n; I.
 DR SEQUENCE 545 AA; 61418 MW; 67B80ACD73C5A9A1 CRC64;
 SQ

Query Match 99.1%; Score 2817; DB 4; Length 545;
 Best Local Similarity 99.4%; Pred. No. 1.6e-216;
 Matches 542; Conservative 1; Mismatches 0; Indels 2; Gaps 2;

QY	1	MLRSKRALPPPLMLLLGPGPLSPGALPRPAQ-ODVVDLDFPQEPHLVSPSFLSVT	58
DB	1	MLRSKRALPPPLMLLLGPGPLSPGALPRPAQ-ODVVDLDFPQEPHLVSPSFLSVT	58
QY	59	VTDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYMS	118
DB	61	VTDANLATDRPFLILGSPKLTARGLSPAYLRFSGTKTDFLIDPKKSTEEERSYMS	120
QY	119	QVQVODICKGSIPEPDVEEKLRLWPYQOULLREHYQKKFKNSTYSRSSVDVLYTFAN	178
DB	121	QVQVODICKGSIPEPDVEEKLRLWPYQOULLREHYQKKFKNSTYSRSSVDVLYTFAN	180
QY	179	CSGLDIFGLNALRTADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFIN	238
DB	181	CSGLDIFGLNALRTADLQNNSSNAQLLDYCSSKGNISWELGNEBNSFLKADIFIN	240
QY	239	QGLGEDYIQLHKLKRSKTFKNAKLYGPDVQGPRAKTKAMLSFLKAGGEVYDSTVMHHYVL	298
DB	241	QGLGEDYIQLHKLKRSKTFKNAKLYGPDVQGPRAKTKAMLSFLKAGGEVYDSTVMHHYVL	300
QY	299	YNGRTATREDPLNDVDLFISSVQKFFOVVESTRPCKTWLGETSSAYGGAAPLSDT	358
DB	301	YNGRTATREDPLNDVDLFISSVQKFFOVVESTRPCKTWLGETSSAYGGAAPLSDT	360
QY	359	FAAGFMWLDKLGSLARMGIEVVMRQVFPAGAGNYHLVDENFDPPLPDYWLSTLLFKLVGKTV	418
DB	361	FAAGFMWLDKLGSLARMGIEVVMRQVFPAGAGNYHLVDENFDPPLPDYWLSTLLFKLVGKTV	420
QY	419	IMAASVQSGSKRRKRLRYLHCTNTDNPYKESGDLTYAINLHVTKYLRPLYPFSNKQVDKYL	478
DB	421	IMAASVQSGSKRRKRLRYLHCTNTDNPYKESGDLTYAINLHVTKYLRPLYPFSNKQVDKYL	480
QY	479	LRLPGLPHGLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPAFSSFFVIIRNAK	538
DB	481	LRLPGLPHGLSKSVQNLGLTLKAVDDOTLPLMEKPLRPSSISGLPAFSSFFVIIRNAK	540
QY	539	VAACT 543	
DB	541	VAACT 545	

RESULT 3
 Q9MUY39 PRELIMINARY; PRT; 545 AA.

AC Q9MUY39; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Heparanase.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 RN NCB1_TaxID=9913;
 RX TISSUE=Placenta;
 RC

RA Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.,
RT "Expression of Heparanase mRNA in Bovine Placenta During Gestation."
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL/AF281160; AF281301.2;
DR InterPro: IPR005199; Glyco_hydro_79n.
DR Pfam: PF03662; Glyco_hydro_79n; 1.
SQ SEQUENCE 545 AA; 61076 MW; FACABDFD855B933 CRC64;

Query Match 80.3%; Score 2282; DB 6; Length 545;
Best Local Similarity 79.8%; Pred. No. 9.5e-174;
Matches 435; Conservative 35; Mismatches 73; Indels 2; Gaps 1;

DB 1 MLARSRPALPPPLML-LLGPGPLSPALPPPAQOVVDLDFTOBRLHVSFSL 58
1 MLACRKGALPPLLLPLGPPGPPGPPPAAPADDAEELFTEERLHLVSPAFSL 60
59 VTIDANLATPRLPILLGSPKRLTARGSPAYLRFGTDTDLFPDKKESTFEERSY 118
61 PTLIDANLATPRLPILLGSPKRLTARGSPAYLRFGTDTDLFPDKKESTFEERSY 120
DB 119 QSGVNDICCKGSIIPDVEEKLRLMPYOBOLLRHYOKKFNSTYSRSVDVLTYPAN 178
121 LSGSNDICCKGSIIPDVEEKLRLMPYOBOLLRHYOKKFNSTYSRSVDVLTYPAN 180
DB 179 CSGDLIFGALNLRADLQWSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFIN 238
181 CSGDLNLFVGNALNLRADLQWSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFIN 240
DB 239 GSGDLGRIYOLNLRKSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSYMHY 298
241 GSGDLGRIYOLNLRKSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSYMHY 300
DB 299 YLNGRTATREDPLNPVLDLFISSVQKVPQVVESTPRGKRWLGERTSSAYGGAPLLSDT 358
301 YVNGRIATKEDPLNPVLDLFISSVQKVPQVVESTPRGKRWLGERTSSAYGGAPLLSDT 360
DB 359 PAAGFMWLDLGLSARNGIEVVRQVFFGAGNHLVDENEDPLDYWLSLFFKLVGTGY 418
361 PAAGFMWLDLGLSARNGIEVVRQVFFGAGNHLVDENEDPLDYWLSLFFKLVGTGY 420
DB 419 LMSVGSRRKRLRVYLHCTNTDNPRIKKEGDLTYAINTLNVTKYLRPVPFSNKVDKY 478
421 LMSVGSRRKRLRVYLHCTNTDNPRIKKEGDLTYAINTLNVTKYLRPVPFSNKVDKY 480
DB 479 LRLPLGPHGLSLKSVQVNGTLTKVNDQDTLPPLMEKELRPSSIGLPAFSYFFVIRNAK 538
481 LRLPLGPHGLSLKSVQVNGTLTKVNDQDTLPPLMEKELRPSSIGLPAFSYFFVIRNAK 540
DB 539 VAACT 543
541 VAACT 545

RESULT 4
08K3 PRELIMINARY; PRT; 535 AA.

AC 08K3;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Heparanase.
OS Mus musculus (Mouse).
OC Mammalia; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RA Maio H.-O., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RT "Cloning, expression, and purification of mouse heparanase."
RL Protein Expr. Purif. 0:0-0(2002).

RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002)
DR EMBL/AY077467; AAL76083.1;
DR EMBL/AK040471; BAC30600.1;
DR MGD; MGI:1343124; Hsae.
DR InterPro: IPR005199; Glyco_hydro_79n.
DR Pfam: PF03662; Glyco_hydro_79n; 1.
SQ SEQUENCE 535 AA; 60065 MW; 6873A8302FB8A0DF CRC64;

Query Match 75.7%; Score 2150; DB 11; Length 535;
Best Local Similarity 76.6%; Pred. No. 3.3e-163;
Matches 407; Conservative 51; Mismatches 73; Indels 0; Gaps 0;

DB 13 LMLLIGPLSPALPPPAQOVVDLDFTOBRLHVSFSLVTIDANLATDPRFL 72
5 LLMWLMGPGLGALAQAPGATPDVVDLEFYTKRPLRSVSPSLITIDSLATDPRFL 64
DB 73 ILLGSPKRLTARGSPAYLRFGTDTDLFPDKKESTFEERSYVNDICCKGSI 132
65 TELGSPRLPALRGLSPAYLRFGTDTDLFPDKKESTFEERSYVNDICCKGSI 124
DB 133 PDVEEKLRLMPYOBOLLRHYOKKFNSTYSRSVDVLTYPANCSGLDIFGALN 192
125 SAAVLRKQVEMPQELLRHQYOKKFNSTYSRSVDVLTYPANCSGLDIFGALN 184
DB 193 KPADLQWSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFINGSLGDIYOLN 252
185 RFPDLKWNSSNAQLLDYSSKGINISWELGNEPNSFLKADIIFINGSLGDIYOLN 244
DB 253 LRKSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSYMHYVYNGRTATREDPL 312
245 LRKSTFKNAKLYGPDVGOPRRKAKMLKSLKAGGEVIDSYMHYVYNGRTATREDPL 304
DB 313 PVLDFIISVQKVPQVVESTPRGKRWLGERTSSAYGGAPLLSDTFAAGFMWLDLGLS 372
305 SVLDFIISVQKVPQVVESTPRGKRWLGERTSSAYGGAPLLSDTFAAGFMWLDLGLS 364
DB 373 ARNGIEVVRQVFFGAGNHLVDENEDPLDYWLSLFFKLVGTGYLMSVGSRRKRL 432
365 ARNGIEVVRQVFFGAGNHLVDENEDPLDYWLSLFFKLVGTGYLMSVGSRRKRL 424
DB 433 VYLHCTNTDNPRIKKEGDLTYAINTLNVTKYLRPVPFSNKVDKYLRPLGPHGLSLK 492
425 VYLHCTNTDNPRIKKEGDLTYAINTLNVTKYLRPVPFSNKVDKYLRPLGPHGLSLK 484
DB 493 VOLNGLTKVNDQDTLPPLMEKELRPSSIGLPAFSYFFVIRNAKVAACI 543
485 VOLNGLTKVNDQDTLPPLMEKELRPSSIGLPAFSYFFVIRNAKVAACI 535

RESULT 5
09QZFB PRELIMINARY; PRT; 536 AA.

AC 09QZFB;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Heparanase.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Podjany K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;

RT "Heparanase from parathyroid cell line";

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL, AF14967; AAF04563.1; -

DR InterPro: IPR005199; Glyco_hydro_79N.

DR Pfam: PF03662; Glyco_hydro_79n; 1.

SO SEQUENCE 536 AA; 60569 MW; 6208B1FD9E28421 CRC64;

Query Match 74.7%; Score 2123; DB 11; Length 536;

Best Local Similarity 75.7%; Pred. No. 4.7e-161; Matches 405; Conservative 51; Mismatches 79; Indels 0; Gaps 0;

```

QY 9 LPPILLLLGPIPLSPGALPRPAQADVDLDPFOEPLHLVSPSFLSVTIDANLATDPRFL 68
DB 2 LRPILLMLMGRLLALNOSTPAGTAFTKDVLDLERYTRQLFQSSPSFLSTTIDASTLTD 61
QY 69 PRFLILGSPKRLTARGLSPAYLRFQGTCTDPLIFDPKKESTFEERSYMQSQVNDICK 128
DB 62 PRFLTFGLSPRLRALRGLSPAYLRFQGTCTDPLIFDPKKESTFEERSYMQSQVNDICK 121
QY 129 YGSLPPVVEKRLRLEMPYQGLLRREHYOKKFKSTYSRSSVDVLTFRANCSGLDILFGL 188
DB 122 SERVSADVLRLQMEWPQGLLRREHYOKKFKSTYSRSSVDVLTFRANCSGLDILFGL 181
QY 189 NALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKAADIFINGSQLGEEDYIQ 248
DB 182 NALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKAADIFINGSQLGEEDYIQ 241
QY 249 LHLKLRKSTFKNAKLYGPDVGPQRRKTAQMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 308
DB 242 LHLKLRKSTFKNAKLYGPDVGPQRRKTAQMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 301
QY 309 DELNPVLDLFISSVQKVFQVVESTRPGKRWLGETSSAYGGAPLSDTFAAGFWMLDK 368
DB 302 DELNPVLDLFISSVQKLVTKKEMTPGKRWLGETSSAYGGAPLSDTFAAGFWMLDK 361
QY 369 LGLSARWGLEIYVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVQSGSKR 428
DB 362 LGLSARWGLEIYVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVQSGSKR 421
QY 429 RKLRYVLTCTNDNPKYKEDGLTYALINLHNTKYRLPYPPSKOVDRYLLRPLGPHGL 488
DB 422 SKLRVYLCTNVVHPRYREGDLTYALINLHNTKYRLPYPPSKOVDRYLLRPLGPHGL 481
QY 489 LSKSVQNLGLTKAVDDOGLPLMEKPLRPSSSLGLPARFSFVIRNAKAIACI 543
DB 482 LSKSVQNLGLTKAVDDOGLPLMEKPLRPSSSLGLPARFSFVIRNAKAIACI 536

```

RESULT 6

```

ID 090YK5 PRELIMINARY; PRT; 523 AA.
AC 090YK5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Heparanase.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11387326;
RA Goldsmith O., Zcharia E., Alingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michael I., Pecker I., Mitrani E., Vlodavsky I.;
RT "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT Determined by Their Signal Peptide Sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
DR EMBL, AY037007; AAK82648.1; -
DR InterPro: IPR005199; Glyco_hydro_79N.
DR Pfam: PF03662; Glyco_hydro_79n; 1.
SO SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

```

Query Match 57.9%; Score 1645.5; DB 13; Length 523;

Best Local Similarity 60.2%; Pred. No. 6.7e-123; Matches 320; Conservative 87; Mismatches 114; Indels 11; Gaps 3;

```

QY 13 LMLLLGPIPLSPGALPRPAQADVDLDPFOEPLHLVSPSFLSVTIDANLATDPRFL 72
DB 2 LVLLLVLLAVLP-----RTTAEQLGIREYIGAVSPFLSTIDASTLARPFRV 52
QY 73 ILGSKPLKTLARGLSPAYLRFQGTCTDPLIFDPKKESTFEERSYMQSQVNDICKYGS 132
DB 53 ALLRHKLHTLASGLSPGLRFQGTCTDPLIFDPKKESTFEERSYMQSQVNDICKYGS 111
QY 133 PPVEKRLRLEMPYQGLLRREHYOKKFKSTYSRSSVDVLTFRANCSGLDILFGLNALL 192
DB 112 FAVPPLLLTQWLOELKLLAEHSWKHKNTTTRSTLILHTFASSGFRVLFGNALL 171
QY 193 RTADLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKAADIFINGSQLGEEDYIQHLK 252
DB 172 RRGGLQWSSNAQLLDYCSSKGYNISWEIGNEPNSFLKKAADIFINGSQLGEEDYIQHLK 231
QY 253 L-RKSTFKNAKLYGPDVGPQRRKTAQMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 311
DB 232 LSGHPLRYRAELYGPDVGPQRRKTAQMLKSFLLKAGEVIDSVTHHYLYNGRTATRE 291
QY 312 NPDVLDLFISSVQKVFQVVESTRPGKRWLGETSSAYGGAPLSDTFAAGFWMLDKLG 371
DB 292 SPVLDLFISSVQKVFQVVESTRPGKRWLGETSSAYGGAPLSDTFAAGFWMLDKLG 351
QY 372 SARWGLEIYVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVQSGSKRKL 431
DB 352 AARGLIDVNRQVFFGAGNYHLDENFDPLPDYMLSLFKLVGTGYTLMAVQSGSKRKL 411
QY 432 RYVLTCTNDNPKYKEDGLTYALINLHNTKYRLPYPPSKOVDRYLLRPLGPHGLSK 491
DB 412 RYVLTCTNDNPKYKEDGLTYALINLHNTKYRLPYPPSKOVDRYLLRPLGPHGLSK 471
QY 492 SYVNLGLTKAVDDOGLPLMEKPLRPSSSLGLPARFSFVIRNAKAIACI 543
DB 472 EYVNLGLTKAVDDOGLPLMEKPLRPSSSLGLPARFSFVIRNAKAIACI 523

```

RESULT 7

```

ID 09HB37 PRELIMINARY; PRT; 592 AA.
AC 09HB37;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Heparanase-like protein HPAC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hicken M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL, AF282887; AAG23423.1; -
DR InterPro: IPR005199; Glyco_hydro_79N.
DR Pfam: PF03662; Glyco_hydro_79n; 1.
SO SEQUENCE 592 AA; 66580 MW; 95C384DA9A74258E CRC64;

```

Query Match 40.6%; Score 1154.5; DB 4; Length 592;

Best Local Similarity 43.6%; Pred. No. 1.5e-83; Matches 250; Conservative 82; Mismatches 189; Indels 53; Gaps 9;

```

QY 20 PLGLSPGAL-----PRPA-----QADVDLDPFOEPLHLVSPS 55
DB 1 LPLGLSPGAL-----PRPA-----QADVDLDPFOEPLHLVSPS 55

```

```

Db      18  PPAACIAGATLALHLHLSLSSOAGDRRLPVDRAAGLKEKTLILLDSTKNPVTNEN 77
Qy      56  FLVSTIDANLADPRFLILSSPKRTLRGLSPAYLRFGTGTDLIF---DPKEST 111
Db      78  FLSLQDPSIIHD-CWLDPLSSKRLVTLARGLSPALRFGKRTDPLQFONLRNPAKSRG 136
Qy      112  FEERSYMSQVNOI-----CKYGSIPPDVEKRLKLEMPYOEOL-LIREHYOK 158
Db      137  GGPDDYILKNYEDDIIVRSVDALDKQCKIAQ-HPDVMLELQREKAQOMHLVLEKEPSN 195
Qy      159  KFKNSTYSSSDVLYTTPANCGLDIFGLNALRTADLQNNSSNAQLLDYCSSKGYNI 218
Db      196  TYSNLIITARSLDKLYNPAQCGSLHLIFALNALRRNPNNSSSALSLKYSKKYNI 255
Qy      219  SWEIGNEPNSFLKADIFINGSQGEDYIQLHKLIRK-STFKNAKLYGPDVGQPRRTAK 277
Db      256  SWEIGNEPNNRTYMGHRAVNGSOLGKDYIQLKSLQPIRISASLYGPNIGRPRKNVIA 315
Qy      278  MLKSFLLKAGGEVIDSYMHYLLNGRTATREDPLNPDVDFISSVOKVFOVVESTRPGK 337
Db      316  LLDGPMKAVGSTVDAVWQHOCYIDGRVVKVMDPLKTRLLDLSQIRKIQKVNTYTPGK 375
Qy      338  KVMIGETSSAYGCGAPLSDTFAAGFMWLDKGLSARMGIEVVRQVFFGAGNYHLVDEN 397
Db      376  KIMLEGVTTTSGAGTNLSDSYAAGFLMNTLGMLANGIDIVIRHSFFDHGNYHLVDON 435
Qy      398  FDPPLDYLILFLFKLVGTKVLMAVSGSKR-----KLAVYLHCTNTDNPYKRG 448
Db      436  FNPPLDYLILFLFKLVGTKVLMAVSGSKR-----KLAVYLHCTNTDNPYKRG 495
Qy      449  DLTITAINLHNTKYLRLPYFPNSKNQYDKYLLRPLGPHGLSKSVQNLGLTKWDDOTL 508
Db      496  SITPLTINLHRSRKKIKIAGTLRDLKLVHQYLLQPYQGEGLSKSVQNLGLTKWDDOTL 555
Qy      509  PLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542
Db      556  PELKRPRLRAGRTLVIPVTMGFFVYKVNALAC 589

```

RESULT 8

```

ID      08WMQ2  PRELIMINARY;  PRT;  592 AA.
AC      08WMQ2;
DT      01-MAR-2002 (TREMBlrel. 20, Created)
DT      01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE      Heparanase 2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RA      Peesegue Salontas B.J.O.P.S.;
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN      (2)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RA      Legoux P., Legoux R., O'Brien D., Salome M.;
RL      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AJ299719; CAC82491.1;
DR      Genew; HGNC:18374; HPSB2.
DR      InterPro; IPR005199; Glyco_hydro_79N.
DR      Pfam; PF03662; Glyco_hydro_79n; I.
SQ      SEQUENCE 592 AA; 66520 MW; 9478641FEACD5588 CRC64;

```

```

Query Match      40.3%; Score 1146.5; DB 4; Length 592;
Best Local Similarity 43.4%; Pred. No. 6.3e-83;
Matches 249; Conservative 82; Mismatches 190; Indels 53; Gaps 9;
20 PLGRLSPAL-----PRPA-----QAQVVDIDFTQEPHLYSS 55

```

```

Db      18  PPAACIAGATLALHLHLSLSSOAGDRRLPVDRAAGLKEKTLILLDSTKNPVTNEN 77
Qy      56  FLVSTIDANLADPRFLILSSPKRTLRGLSPAYLRFGTGTDLIF---DPKEST 111
Db      78  FLSLQDPSIIHD-CWLDPLSSKRLVTLARGLSPALRFGKRTDPLQFONLRNPAKSRG 136
Qy      112  FEERSYMSQVNOI-----CKYGSIPPDVEKRLKLEMPYOEOL-LIREHYOK 158
Db      137  GGPDDYILKNYEDDIIVRSVDALDKQCKIAQ-HPDVMLELQREKAQOMHLVLEKEPSN 195
Qy      159  KFKNSTYSSSDVLYTTPANCGLDIFGLNALRTADLQNNSSNAQLLDYCSSKGYNI 218
Db      196  TYSNLIITARSLDKLYNPAQCGSLHLIFALNALRRNPNNSSSALSLKYSKKYNI 255
Qy      219  SWEIGNEPNSFLKADIFINGSQGEDYIQLHKLIRK-STFKNAKLYGPDVGQPRRTAK 277
Db      256  SWEIGNEPNNRTYMGHRAVNGSOLGKDYIQLKSLQPIRISASLYGPNIGRPRKNVIA 315
Qy      278  MLKSFLLKAGGEVIDSYMHYLLNGRTATREDPLNPDVDFISSVOKVFOVVESTRPGK 337
Db      316  LLDGPMKAVGSTVDAVWQHOCYIDGRVVKVMDPLKTRLLDLSQIRKIQKVNTYTPGK 375
Qy      338  KVMIGETSSAYGCGAPLSDTFAAGFMWLDKGLSARMGIEVVRQVFFGAGNYHLVDEN 397
Db      376  KIMLEGVTTTSGAGTNLSDSYAAGFLMNTLGMLANGIDIVIRHSFFDHGNYHLVDON 435
Qy      398  FDPPLDYLILFLFKLVGTKVLMAVSGSKR-----KLAVYLHCTNTDNPYKRG 448
Db      436  FNPPLDYLILFLFKLVGTKVLMAVSGSKR-----KLAVYLHCTNTDNPYKRG 495
Qy      449  DLTITAINLHNTKYLRLPYFPNSKNQYDKYLLRPLGPHGLSKSVQNLGLTKWDDOTL 508
Db      496  SITPLTINLHRSRKKIKIAGTLRDLKLVHQYLLQPYQGEGLSKSVQNLGLTKWDDOTL 555
Qy      509  PLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 542
Db      556  PELKRPRLRAGRTLVIPVTMGFFVYKVNALAC 589

```

RESULT 9

```

ID      08WMQ1  PRELIMINARY;  PRT;  548 AA.
AC      08WMQ1;
DT      01-MAR-2002 (TREMBlrel. 20, Created)
DT      01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE      Heparanase 3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RA      Peesegue Salontas B.J.O.P.S.;
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN      (2)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RA      Legoux P., Legoux R., O'Brien D., Salome M.;
RL      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AJ299720; CAC82492.1;
DR      InterPro; IPR005199; Glyco_hydro_79N.
DR      Pfam; PF03662; Glyco_hydro_79n; I.
SQ      SEQUENCE 548 AA; 61771 MW; B8986303FC7A60A CRC64;

```

```

Query Match      35.7%; Score 1015.5; DB 4; Length 548;
Best Local Similarity 41.8%; Pred. No. 1.7e-72;
Matches 224; Conservative 79; Mismatches 180; Indels 53; Gaps 9;
8 ALPPLMLLLGCPPLSPAL-----PRPA-----QAQVVDIDF 43
6 APEEMLSNSRPPACIAGATLALHLHLSLSSOAGDRRLPVDRAAGLKEKTLILLD 65

```

QY 44 PROBPLHVSPEFLSTIDANLATDPRFLILGSPKRTLAGSPAYLFGCTKIDFLI 103
 DB 66 STKNPRTYNNENFLSLQIDPSIHD-GMLDPLSSKRLVTLNGLSPAFKFGKRTDFO 124
 QY 104 F-----DPKKESTFEERSYVQOVNODI-----CKYGSIPDVEEKLRLWPYQ 147
 DB 125 FGNLNPAPKSRGPGPDYLLKNYEDDIIVASDVALDKQCKCIAQ--HPDVMELQREKAAQ 183
 QY 148 EOL-LIREHYOKKFKKSTYSRSSVDLYTFPANCSGIDLIFGNALLRTADLQWSSNAOL 206
 DB 184 MELVLLKEQPSYTSYSLILITARSLLDLYNSADCSGHLIPALNALRRNNNSWSSALS 243
 QY 207 LIDYSSSKGYNISWELGNEPNSFLKADIFINGSOLEDYIOLHKLK--STFKNAKYQ 265
 DB 244 LKTSASKKTYNISWELGNEPNSFLKADIFINGSOLEDYIOLHKLK--STFKNAKYQ 303
 QY 266 PDVGPRRRTAKMLKSLKAGGEVIDSVTMHYYLNGRTATREDFLNPVDLDFISSVOK 325
 DB 304 PIGRPRKRVIALLOGFVKVAGSTYDAVMOHCYIDGRVVKVMDPLKTRLDLTLSDQIRK 363
 QY 326 VPOVVESTRPGKRWLGSTSSAYGGAPLLSDTPAAGFMWDLGLSARNGIEVWROVYF 385
 DB 364 IOKVNTYTPGKKIMLEGVVTTSAGTNNLSDSYAAGFLMLNTLGLANOGIDVYIRHSF 423
 QY 366 PCAGNYHLVDENFDPLPDYMLSLPFKLVGTKYLMASVOGSKRR-----KLRYVYLH 436
 DB 424 FPHGTHLVDQNFPLPDYMLSLPFKLVGTKYLMASVOGSKRR-----KLRYVYLH 483
 QY 437 CTNTDNPARYEGDLYTVAINLHNTYKRLPYFPNSKQVQDKYLLRPLGPHGLSKS 492
 DB 484 CTNNHNNYVVRGSITLFIINLHRSKKIKLAGTLRDLKHQYLLQYGGGLSKSKT 539

RESULT 10

Q9HB38 PRELIMINARY; PRT; 534 AA.
 ID AC Q9HB38
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DB Hepatanae-like protein HPA2b.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20483645; PubMed=11027606;
 RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
 RA Hircok M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.,
 RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
 RT Hepatanae Family Member."
 RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
 DR EMBL: AF282886; AAC23422.1;
 DR InterPro: IPR005199; Glyco_hydro_79N.
 DR Pfam: PF03662; Glyco_hydro_79N.
 QY SEQUENCE 534 AA; 60063 MW; C3DSEB900CB38C4 CRC64;

Query Match 33.0%; Score 936.5; DB 4; Length 534;
 Best Local Similarity 37.8%; Pred. No. 3.3e-66;
 Matches 217; Conservative 78; Mismatches 168; Indels 111; Gaps 12;

QY 20 PLGSPSPAL-----PRPA-----QAQDVVDLDFTOEPFLHVS 55
 DB 18 PACIAPAGLIALALHLSLSSQAGRRPLVDRAAGLKEKTLILDVSTKMPVTVNEN 77
 QY 56 FLSTVDANLATDPRFLILGSPKRTLAGSPAYLFGCTKIDFLI-----DPKKEST 111
 DB 78 FLSTQIDPSIHD-GMLDPLSSKRLVTLNGLSPAFKFGKRTDFOQNLNPAKSR 136
 QY 112 FEERSYVQOVNODI-----CKYGSIPDVEEKLRLWPYQOL-LIREHYOK 158

DB 137 GPGPDYLLKNYEDDIIVASDVALDKQCKCIAQ--HPDVMELQREKAAQMLVLLKEQF-- 193
 QY 159 KFKSTYSRSSVDLYTFPANCSSGLDLIFGNALLRTADLQWSSNAOLLDYSSKGYNI 218
 DB 194 ---SNTYS-----NLIT----- 202
 QY 219 SWEIGNEPNSFLKADIFINGSOLEDYIOLHKLK--STFKNAKYGPDVGPRRTAK 277
 DB 203 ---TEPNYNTMNGRAVNSQGLKDIOLKSLLOPIRISBSASLVGPNIGRRKQVIA 257
 QY 278 MLKSPFLKAGGEVIDSVTMHYYLNGRTATREDFLNPVDLDFISSVOKVPOVVESTRPGK 337
 DB 258 LIDGFMKVAAGSTYDAVMOHCYIDGRVVKVMDPLKTRLDLTLSDQIRKIQVNTYTPGK 317
 QY 338 KWLIGETSSAYGGAPLLSDTPAAGFMWDLGLSARNGIEVWROVYFPGAGNYLVDEN 397
 DB 318 KIMLGVVTSAGTNNLSDSYAAGFLMLNTLGLANOGIDVYIRHSFFDHGYNHLDON 377
 QY 398 FDPPLDYMLSLPFKLVGTKYLMASVOGSKRR-----KLRYVYLHCTNTDNPARYEG 448
 DB 378 FNPPLDYMLSLYKRLIGPKYLAHVAGLQKRPGRVIRDKRLIYAHCTNNHNNYVRG 437
 QY 449 DLTVAINLHNTYKRLPYFPNSKQVQDKYLLRPLGPHGLSKSVOLNGLTLKAVDDOTL 508
 DB 438 SITLFIINLHRSKKIKLAGTLRDLKHQYLLQYGGGLSKSVOLNGLTLKAVDDOTL 497
 QY 509 PLMEKPLRPGSSIGLPAFSYFVIRAKVLAAC 542
 DB 498 PELKRPFLRAGRTLVIPVTWGFVAKVNALAC 531

RESULT 11

Q9HB39 PRELIMINARY; PRT; 480 AA.
 ID AC Q9HB39
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DB Hepatanae-like protein HPA2a.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20483645; PubMed=11027606;
 RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
 RA Hircok M., Patel S., Barry E., Stuberfield C., Terrett J., Page M.,
 RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
 RT Hepatanae Family Member."
 RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
 DR EMBL: AF282885; AAC23422.1;
 DR InterPro: IPR005199; Glyco_hydro_79N.
 DR Pfam: PF03662; Glyco_hydro_79N.
 QY SEQUENCE 480 AA; 53900 MW; F75F89F67AC1FF83 CRC64;

Query Match 31.6%; Score 897.5; DB 4; Length 480;
 Best Local Similarity 36.0%; Pred. No. 3.7e-63;
 Matches 202; Conservative 74; Mismatches 146; Indels 139; Gaps 9;

QY 20 PLGSPSPAL-----PRPA-----QAQDVVDLDFTOEPFLHVS 55
 DB 18 PACIAPAGLIALALHLSLSSQAGRRPLVDRAAGLKEKTLILDVSTKMPVTVNEN 77
 QY 56 FLSTVDANLATDPRFLILGSPKRTLAGSPAYLFGCTKIDFLI-----DPKKEST 111
 DB 78 FLSTQIDPSIHD-GMLDPLSSKRLVTLNGLSPAFKFGKRTDFOQNLNPAKSR 135
 QY 112 FEERSYVQOVNODICKYGSIPDVEEKLRLWPYQOL-LIREHYOKKFKKSTYSRSSVD 171
 DB 136 -----GPGPD-----YLLKYE----- 148
 QY 172 VLYTFPANCSSGLDLIFGNALLRTADLQWSSNAOLLDYSSKGYNISWELGNEPNSFLK 231

DB 149 -----DEPNNRT 156
 QY 232 KADIFNGSOLGADYIOLHKLARK-STFNNAKLYGPDVGOPRRKTAAMLSFLKAGGEVI 290
 DB 157 MGRGAVNSQGLQKOVITOLKSLQIPRIYSRSLVGPRIKPRKAVIALLDGMKAGSTV 216
 QY 231 DSVTHHTYLLNGRTYREDPLNDVLDIFISSVQKVFQVVESTTRGKVTGCTSSAYG 350
 DB 217 DAVTQHCYIDRVRKVDVDFLTRLDLDLSDQIRKIOKAVNYTTEGKIMEGVVTTSAG 276
 QY 351 GADLSDPFAAGFMWLDGLSABMGIEVMRQVFPAGNHYLVDENDPLPDVWLSLTF 410
 DB 277 GNNLSDSYPAAGFLMNTLGMLANOGIDVIRHSFDRGYNHLVDQNFNPDPYWLSTLY 336
 QY 411 KKLVTCTVLAASVQSGSKR-----KLRYLHCTNTDNPYKSGDLTYALINAHVT 461
 DB 337 KRLIGPKYLAHVAVGLQKRRGRVIRKLRITACHTNHNHNYRGSITLFIILHRSR 396
 QY 462 KYLRIPYPSNQVDKYLRLPGLPHGLLSKSVQNLGLTKVDDOTLPLMEKPLRPGSS 521
 DB 397 KKIKLAGTIRBKLYVQYLLQPYGQGLSKSVQNLGCPVWVDDGTLBELKPRPLRAGRT 456
 QY 522 LGPAFSYSPFYIRAKYAAAC 542
 DB 457 LVIPVTMGFFVVKVNAALAC 477

RESULT 12

ID 08T108 PRELIMINARY: PRT: 515 AA.
 AC 08T108:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 GN Heparanase-like protein.
 CN BHEPRA.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditryta; Bombycoidea;
 OC Bombycidae; Bombyx.
 CX NCBI_TaxId=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=50; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 "Genomic sequence of 320kb containing a kettin orthologue on the 2
 RT chromosome in Bombyx mori."
 RT Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=50; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 "Genomic sequence of 320kb containing a kettin orthologue on the 2
 RT chromosome in Bombyx mori."
 RT Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.
 RL EMBL; AB079860; BAB8519.1;
 DR EMBL; AB090307; BAB10612.1;
 DR InterPro; IPR005199; Glyco_hydro_79N.
 DR InterPro; IPR006895; zf_Sec23_Sec24.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 DR Pfam; PF04810; zf_Sec23_Sec24; 1.
 DR SEQUENCE 515 AA; 59769 MW; FB8100ABEEDDADB CRC64;

Query Match 24.5%; Score 636; DB 5; Length 515;
 Best Local Similarity 35.1%; Pred. No. 5.3e-47;
 Matches 183; Conservative 83; Mismatches 182; Indels 74; Gaps 18;

QY 46 GEPFLVSEFSLVITDANLADPRFLILSGPKLRTLAGSPAYIRGCTKTFLLFD 105
 DB 42 QBDIKLISDFLSFGID-TIIEENYRINYSOTRLRELAALSPARLIRGTMSSRLIF- 99

QY 106 PKKSTFEERSYNQSVQVODICKYGISIPDVEEKLRLMEPYQOBLLREHYOKKFNSTY 165
 DB 100 -SKENI-----PISCHNCSTYKYSKISCO--LLEKPC-----KHHKFLPFPIIM 140
 QY 166 SRSSVDVLYTFANSGDGLIFGLNALRLTADLDWNSNAOLLDDYSSKGYNISMEIGNE 225
 DB 141 TGNWNOINDFCRKTINLKLFLSLNAALRD-NHGNNEKRAELIFSKHKQVAILMOGLNE 199
 QY 226 PMSFLKADIFINQSGEDYIOLHKLARKSTFNNAKLYGPDVGOP--RRKTAAMLSKF 282
 DB 200 PMSQVHFNESVTFQILAKQPEKRLKLNHNGYHSLIVGDTTRPOPHRECKLYMIEF 259
 QY 283 LKAGGEYIDSTVHTYLLNGRTATREDPLNDVLDIFISSVQKVFQVVESTTRGKVTGCTSSAYG 340
 DB 260 LGNSHYINRSMQVYINSATLDEDFWMBETDLDL--RQIETQMONQKTKYKNTLPMW 316
 QY 341 LGFTSSAYGGAPLSDTPAAGFMWLDGLSABMGIEVMRQVFPAGNHYLVDENDPL 400
 DB 317 LSETSSSYGGAPLSDTPAAGFMWLDGLSABMGIEVMRQVFPAGNHYLVDENDPL 375
 QY 401 LPDYWLSLFPKQVGTCTVLAASVQSGSKRRLKRYLHCTNTDNPYK--GDLTYALIN- 456
 DB 376 LPDMWISVLYKLVGNKVL--QVQCNCSSRFQRLYIHCNTR--KTYNDTSAYVLYGVNLB 430
 QY 457 -----LHN---VTKYLRIPYPSNQVDKYLRLPGLPHGLLSKSVQNLGLTKVDDOTLPLMEKPLRPGSS 521
 DB 431 MAKARFPLNGTALHGDLLIHEYI-ISAPSNRK-----SKTILNGMPL 474
 QY 501 KMWDDQTLPLMEKPLRPGSSLGPAFSYSPFYIRAKYAAAC 542
 DB 475 YY--ESNLHNRPNIRHRYGRYVSLPPYSIGFWIKTSITVC 514

RESULT 13

ID 09SDA1 PRELIMINARY: PRT: 521 AA.
 AC 09SDA1:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical 57.8 kDa protein.
 GN Flg24.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eupsidia II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Beyer M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
 RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC EU Arabidopsis sequencing project;
 RA Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 RL EMBL; AL133421; CAB62595.1;
 DR InterPro; IPR005199; Glyco_hydro_79N.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hypothetical protein.
 DR SEQUENCE 521 AA; 57831 MW; 07D8664AAB305CC2 CRC64;

Query Match 14.6%; Score 416; DB 10; Length 521;
 Best Local Similarity 29.2%; Pred. No. 1.3e-24;
 Matches 154; Conservative 68; Mismatches 184; Indels 122; Gaps 24;

QY 75 LGSPLRTLAGLSPAYIRFGCTKTFLLFPKKESTFEERSYNQSVQVODICKYGISIP 134
 DB 55 LTRPLTATATAPFPRIRIGGSLQDVYIVGVNLT-----PCR----- 94

QY 135 DVEBKRLNMPYQEBOLLREHYOKKFKNS---TYSRSSV---DVLTYFANGSGDLIF 186
 Db -----PFOKM-----NSGLFSGSKCLHMKRMDLNSFLTAGVAVTF 132
 QY 187 GINALLRTADLQ-----MNSSNAQLLDYCSSKGYNI-SWELGNEPNFLKKADIFIN 238
 Db GINALRGHKLKRGKAWGAMDHINTODPLNNTVSGVYIDSWERFENLSG--SGVGASVS 190
 QY 239 GSQLEDYIQLHLKLRSTFKNAKLYGPDVGP-----RRKTAUKLSFLKAGGEVIDSV 293
 Db 191 AELYGKDLVLKQVINK-VYKNSWLHKPLIIVAPGPFYEQMYTKLLET---SGPSVADV 246
 QY 294 TWHYYLNGRT--ATREDFLNPDVLDIFISSVQKVF---QVESTRPCKVWLGSETSA 347
 Db 247 THHIVNLGSGNDPALVKKIMDPS---YLSQVSKTFKDVNQTIOEHGPMASPMVGEISGA 302
 QY 348 YGGGAPLLSDPFAAGFMWLDLGLSABMGIEVWQVFFGAGNTHLVD--NEDPLPDYWL 406
 Db 303 YNSGGRHVSDTFIDSFYLDQLGMSARHNTKYVCROTLVG--GFYGLLEKGTFFVNPDDYS 361
 QY 407 SLLEKLVGTQVTLVMAVSQSKRKLRYVLIHCTNTDNPYKEGDLTYAINLHNTKYL-- 464
 Db 362 ALLMRLMKGKVLAVQTDGP--QLRYVAHCSK-----GRAGVTLLILNLSQSDFTVS 413
 QY 465 -----RLPYPS--NKQVDKYLRLP--LGPBG--LLSKSVQL 495
 Db 414 VSNGINVVLNABSRKKSLDLTLKRPFSWIGSKASDGYLNEEYHLTPENGVLRSKTVWL 473
 QY 496 NGTLTKWDDQTLPLMEKPLRP-GSSGLPAFSYFPVIRNAKVAAC 542
 Db 474 NKSILKPTATGDIPLS-LPVLRSVNSPLNVLPLSISFIVLPNPDASAC 520

RESULT 14

QY 09FP10 PRELIMINARY; PRT; 543 AA.
 AC Q9FP10; 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similarity to heparanase.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=97471969; PubMed=9330910;
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asanizu E., Fukami M.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT clones."
 RT DNA Res. 4:215-230(1997).
 DR EMBL; AB005249; BAB09847.1;
 DR InterPro; IPR005199; Glyco_hydro_79N.
 DR InterPro; IPR001254; Ser_protease_tty.
 DR Pfam; PF03662; Glyco_hydro_79n.1.
 DR PROSITE; PS00135; TRYPSIN_SRR.1.
 SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Query Match 14.6%; Score 416; DB 10; Length 543;

Best Local Similarity 29.2%; Pred. No. 1,4e-24; Matches 154; Conservative 68; Mismatches 184; Indels 122; Gaps 24;

QY 75 LGSPLRTIARGLSPAYLRFGTGTDPLIFDPKSTFEERSYWGQVNOIDCKYCSIPP 134
 Db 77 LRPPLTKAIKAKPARIKIGSLQDOVIYDGNLKT-----PCR----- 116
 QY 135 DVEBKRLNMPYQEBOLLREHYOKKFKNS---TYSRSSV---DVLTYFANGSGDLIF 186

Db 117 -----PFOKM-----NSGLFSGSKCLHMKRMDLNSFLTAGVAVTF 154
 QY 187 GINALLRTADLQ-----MNSSNAQLLDYCSSKGYNI-SWELGNEPNFLKKADIFIN 238
 Db 155 GINALRGHKLKRGKAWGAMDHINTODPLNNTVSGVYIDSWERFENLSG--SGVGASVS 212
 QY 239 GSQLEDYIQLHLKLRSTFKNAKLYGPDVGP-----RRKTAUKLSFLKAGGEVIDSV 293
 Db 213 AELYGKDLVLKQVINK-VYKNSWLHKPLIIVAPGPFYEQMYTKLLET---SGPSVADV 268
 QY 294 TWHYYLNGRT--ATREDFLNPDVLDIFISSVQKVF---QVESTRPCKVWLGSETSA 347
 Db 269 THHIVNLGSGNDPALVKKIMDPS---YLSQVSKTFKDVNQTIOEHGPMASPMVGEISGA 324
 QY 348 YGGGAPLLSDPFAAGFMWLDLGLSABMGIEVWQVFFGAGNTHLVD--NEDPLPDYWL 406
 Db 335 YNSGGRHVSDTFIDSFYLDQLGMSARHNTKYVCROTLVG--GFYGLLEKGTFFVNPDDYS 383
 QY 407 SLLEKLVGTQVTLVMAVSQSKRKLRYVLIHCTNTDNPYKEGDLTYAINLHNTKYL-- 464
 Db 364 ALLMRLMKGKVLAVQTDGP--QLRYVAHCSK-----GRAGVTLLILNLSQSDFTVS 435
 QY 465 -----RLPYPS--NKQVDKYLRLP--LGPBG--LLSKSVQL 495
 Db 436 VSNGINVVLNABSRKKSLDLTLKRPFSWIGSKASDGYLNEEYHLTPENGVLRSKTVWL 495
 QY 496 NGTLTKWDDQTLPLMEKPLRP-GSSGLPAFSYFPVIRNAKVAAC 542
 Db 496 NKSILKPTATGDIPLS-LPVLRSVNSPLNVLPLSISFIVLPNPDASAC 542

RESULT 15

QY 08H615 PRELIMINARY; PRT; 544 AA.
 AC Q8H615; 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE OSJUBA0035103.11 protein.
 GN OSJUBA0035103.11.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC
 RT clone:OSJUBA0035103."
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003019; BAC22238.1;
 SQ SEQUENCE 544 AA; 58723 MW; ECD8695F0E226269 CRC64;

Query Match 13.8%; Score 392.5; DB 10; Length 544;

Best Local Similarity 26.8%; Pred. No. 1e-22; Matches 144; Conservative 58; Mismatches 179; Indels 157; Gaps 19;

QY 75 LGSPLRTIARGLSPAYLRFGTGTDPLIFDPKSTFEERSYWGQVNOIDCKYCSIPP 125
 Db 93 LSNKILNAIRAFSPDLKRLIGSLQDKLVYGVGDGCGCAFFVKNVTSMPFGTQ----- 146
 QY 126 ICKYSIIPDVBKRLNMPYQEBOLLREHYOKKFKNSTYSSSDVLYTFANGSGDLIF 185
 Db 147 -----GCLPLH-----RW-----DELNAFFQSGARIV 169
 QY 186 GINALLRTADLQ-----MNSSNAQLLDYCSSKGYNI-SWELGNEPNFLKKADIFIN 237
 Db 170 FGALNLNRPVLPDGSQMGPPDYNTAASLIRYTSKGIKHWELGNE-----L 218
 QY 238 NSQLEDYIQLHLKLRSTFKNAKLYGPDV-----QPRRTAKMLKSLFLKA 285

06:21:55 2003

Db 219 SSGVGC-----TKVADQYADYIALKSLVDVTIYQGNPSKPLVLA PGFPDA 265
QY 286 G--GEVID-----SVTWHHYTLNG--RTATREDFLNPDVLDIFISSVCKVFOVBESE 333
266 GWFTBEVIVKTRPNLINVVTHHTYNLGPVDTLIEKILNPSYLDGKVSFTSNLOGILKSA 325
QY 334 RPKKXWLGETSAYCGGAPLSDTFAAGFMWLDKLSABMGIEVWVRQVFFGAGNYHL 393
Db 326 GTSAYAMVGBSGAIVNSGRHLVTDSEVFSFWFLDQJMSAKYDTKSYCRQSLIG--GNYGL 384
QY 394 VD--ENFDPLPDYWLSLFFKKLVGTVKVLMSVOGSKRRKLRVYLHCTNTDNPRYKBDLTL 452
Db 385 LNKETQPNPDYYSALLMHRIMGTKVL SATFNCT--NMIRTYAHCAK--DSP-----GITL 436
QY 453 YAINL-----HNTKYLRLPYPPSNKQVDKYLRLPLG 484
Db 437 LILNLSGNTTSQVSVTSEGANHTVKKHSRTRHLAGSNREBYHLJAKD----- 486
QY 485 PHGLSKSVOLNGLTLKMDOTLPLMEKPLRPGSSIGLPAFSYSFFVI RNKVAAC 542
Db 487 -GSLQSQVMLNGLRALVADENGELPRLEPVKVDAAOPIAVAPYSIVFAHIHNFAPAC 543

Search completed: October 22, 2003, 20:26:50
Job time : 110 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 20:19:45 ; Search time 86 Seconds
(without alignments)
1002.192 Million cell updates/sec

Title: US-09-759-207-2

Perfect score: 2842
Sequence: 1 MLRSKRALPPPLMLLLGP.....LPARSFPVIRNAKVAACI 543

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2842	100.0	543	AAV02345	A human heparanase
2	2842	100.0	543	AA808849	Amino acid sequenc
3	2842	100.0	543	AAV57590	Human heparanase.
4	2842	100.0	543	AAV52990	Human heparanase p
5	2842	100.0	543	AAV97635	Human heparanase p
6	2842	100.0	543	AB807813	Human heparanase s
7	2842	100.0	592	AAV02346	A human heparanase
8	2842	100.0	592	AA808850	Amino acid sequenc
9	2838	99.9	543	AAV17082	Human heparanase e

10	2838	99.9	543	22	AA86206	Human heparanase 1
11	2838	99.9	543	20	AAV30124	A human protein w/
12	2826	99.1	543	22	AA889361	Human membrane or
13	2817	99.1	545	24	ABP56822	Human heparanase p
14	2764	97.3	530	20	AAV34173	Human pre-prohepar
15	2737	96.3	532	20	AAV17083	Seq ID No: 15 of W
16	2671.5	94.1	527	23	AB807815	Chicken signal pep
17	2146	75.5	535	21	AA808851	A murine heparanas
18	2146	75.5	535	23	AB807811	Mouse heparanase s
19	2123	74.7	536	23	AB807812	Rat heparanase seq
20	1645.5	57.9	523	23	AB807814	Chicken heparanase
21	1614	56.8	380	20	AAV17085	Rat heparanase enz
22	1602	56.4	380	20	AAV17084	Mouse heparanase e
23	1154.5	40.6	532	22	AAU07424	Human heparanase-1
24	1148.5	40.6	532	22	AAV97632	Human heparanase-2
25	1148.5	40.4	532	22	AA881062	Human heparanase-2
26	1147.5	40.4	532	22	AA885215	Heparanase-1 like pr
27	1142.5	40.2	582	23	AAE18326	Human heparanase-2
28	1132.5	39.1	538	22	AAV97633	Human heparanase
29	1106.5	38.9	528	23	AAE18327	Human heparanase-2
30	936.5	33.0	534	22	AA885216	Heparanase-like pr
31	936.5	33.0	534	23	ABP69310	Human polypeptide
32	936.5	33.0	534	23	AAV50337	Human prepro-hepar
33	927.5	32.6	492	22	AA884664	Amino acid sequenc
34	897.5	31.6	480	22	AAU07418	Novel human extrac
35	897.5	31.6	480	22	AA885217	Heparanase-like pr
36	887.5	31.6	480	22	AAV97634	Human heparanase
37	892.5	31.4	470	23	AAE18328	Human heparanase-2
38	891.5	31.4	439	23	AAU07423	Human heparanase-1
39	788	27.7	331	23	AAV50383	Human heparanase I
40	663	23.3	488	22	AA831469	Amino acid sequenc
41	645	22.7	488	22	AA831470	Amino acid sequenc
42	642	22.6	488	22	AA831472	Amino acid sequenc
43	622	21.9	488	22	AA831471	Amino acid sequenc
44	528.5	18.6	214	22	AAV99905	Human excretory re
45	528.5	18.6	214	22	AAV43704	Human bladder anti

ALIGNMENTS

RESULT 1
AAV02345
ID AAV02345 standard; Protein; 543 AA.
XX
AC AAV02345;
XX
XX 09-JUL-1999 (first entry)
XX
DE A human heparanase protein.
XX
XX Heparanase; hp; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; resensitise;
KW atherosclerosis; inflammation; neurodegenerative disease; neuritis;
KW plasma heparin; microtubulastasis; autoimmune lesion; renal failure.
XX
OS Homo sapiens.
XX
XX PD WO911798-A1.
XX PD 11-MAR-1999.
XX PF 31-AUG-1998; 98WO-US17954.
XX PR 02-JUL-1998; 98US-0109386.
XX PR 02-SEP-1997; 97US-0922170.
XX
XX (FRIE/) FRIEDMAN M M.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (INST-) INSIGHT STRATEGY & MARKETING LTD.

Query Match	Best Local Similarity	100.0%	Score 2842	DB 20	Length 543
Matches 543	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1	MLNRSKPALPPPLMLLLGPGPLSPGALPRPAQADVDVDDLPFOEPHLVSPFLSVT	60		
DB	1	MLNRSKPALPPPLMLLLGPGPLSPGALPRPAQADVDVDDLPFOEPHLVSPFLSVT	60		
QY	61	IDANLAPTRPFLIILGSPPLRTLAGLSPAYLRFGSTKTDPLIFDPCKKSTPEERSYVOS	120		
DB	61	IDANLAPTRPFLIILGSPPLRTLAGLSPAYLRFGSTKTDPLIFDPCKKSTPEERSYVOS	120		
QY	121	QVNDICIKYSGIPRPVEBEKRLLEWYEOBOLLREHYOKKFNKSTYSRSSVDLYTFPANC	180		
DB	121	QVNDICIKYSGIPRPVEBEKRLLEWYEOBOLLREHYOKKFNKSTYSRSSVDLYTFPANC	180		
QY	181	GLDLIFGLNALIKRTADLQWNSNAOLLDDYCSKGINISWEIGNEPNSFLKKAADIFINGS	240		
DB	181	GLDLIFGLNALIKRTADLQWNSNAOLLDDYCSKGINISWEIGNEPNSFLKKAADIFINGS	240		
QY	241	OLGEBYIOLHKLIRKSTFPKNAULYGPVDQGPBRKATKMLKSLFKKAGEYIDVTHHYLL	300		
DB	241	OLGEBYIOLHKLIRKSTFPKNAULYGPVDQGPBRKATKMLKSLFKKAGEYIDVTHHYLL	300		
QY	301	NGRTATREDPLNPVDLDFISSVQKVFQVESTTRPGKTKWMLGETSSAYGCGAPLLSDTPA	360		
DB	301	NGRTATREDPLNPVDLDFISSVQKVFQVESTTRPGKTKWMLGETSSAYGCGAPLLSDTPA	360		
QY	361	AGFWMLDKLGISARMGIEVYMRQVFFGAGNYHLVDENFDPLDPYMLSLLFKKLVGKTYLM	420		
DB	361	AGFWMLDKLGISARMGIEVYMRQVFFGAGNYHLVDENFDPLDPYMLSLLFKKLVGKTYLM	420		
QY	421	ASVQSGKRRKRLRVYLHCTNTDNPARKSGDLTYALNLHNTTKYLRPYFNSKQVDDKYL	480		
DB	421	ASVQSGKRRKRLRVYLHCTNTDNPARKSGDLTYALNLHNTTKYLRPYFNSKQVDDKYL	480		
QY	481	RPLPGHGLLSKSVOLNGTLTKWVDDOTLPPLMEKPLRPSSIGLPAFSSYFVIRNAKVA	540		
DB	481	RPLPGHGLLSKSVOLNGTLTKWVDDOTLPPLMEKPLRPSSIGLPAFSSYFVIRNAKVA	540		
QY	541	ACTI 543			
DB	541	ACTI 543			

ID	AA	08849 standard; Protein; 543 AA.
XX	AA	08849;
XX	DT	15-JAN-2001 (first entry)
DE	XX	Amino acid sequence of a human heparanase polypeptide.
XX	XX	Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW	XX	heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW	XX	wound healing; infection; burn; angiogenesis; restenosis;
XX	XX	atherosclerosis; inflammation; neurodegenerative disease;
XX	XX	Geistmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
OS	XX	Homo sapiens.
XX	PN	MO200052178-A1.
XX	PD	08-SEP-2000.
XX	PE	14-FEB-2000; 2000WO-US03542.
XX	PR	01-MAR-1999; 99US-0258892.
XX	PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX	PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PI	PI	(FRIE/) FRIEDMAN M M.
XX	PI	Pecker I, Violdavsky I, Feinstein E;
DR	DR	WPI: 2000-579289/54.
XX	XX	N-PSDB; AAA75051.
PT	PT	New polynucleotides encoding a polypeptide having heparanase activity,
PT	PT	useful in wound healing and in gene therapy, particularly in treating
XX	XX	tumour, inflammation, autoimmunity, neurodegenerative diseases
XX	XX	Claim 22; Fig 1; 152pp; English.
XX	XX	The present sequence represents a human protein with heparanase catalytic
CC	CC	activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC	CC	particularly in treating tumour, inflammation or autoimmunity.
CC	CC	Particularly, the polynucleotide is useful in modulating the
CC	CC	bioavailability of heparin-binding growth factors, cellular responses
CC	CC	to heparin-binding growth factors (e.g. bFGF) and cytokines
CC	CC	(e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins,
CC	CC	cellular susceptibility to certain viral and some bacterial and protozoa
CC	CC	infections, or disintegration of neurodegenerative plaques. The
CC	CC	polynucleotide is also useful in wound healing (e.g. thermal, chemical
CC	CC	or radiation burns), and in the treatment of angiogenesis, restenosis,
CC	CC	atherosclerosis, inflammation, neurodegenerative diseases (Geistmann-
CC	CC	Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
XX	XX	bacterial or protozoa infections.
SQ	SQ	Sequence 543 AA;
Query Match	100.0%;	Score 2842; DB 21; Length 543;
Best Local Similarity	100.0%;	Pred. No. 9.5e-274;
Matches 543;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MURSKPALPPMLMLLGLPGLSPGALPPPAQADVVDDFFQOEPLHUVSPFLSVT 60
DB	1	MLTRSKPALPPMLMLLGLPGLSPGALPPPAQADVVDDFFQOEPLHUVSPFLSVT 60
QY	61	IDANLADPRFLILGSPKLTARGLSPALRFGCTDTDFLIPDKKESFPEESYWOS 120
DB	61	IDANLADPRFLILGSPKLTARGLSPALRFGCTDTDFLIPDKKESFPEESYWOS 120
QY	121	QVNDQICKYSIPPDVEBKRLLEMPYOEQLLREHYOKKFNKSTYSRSVDVLYTFANC 180
DB	121	QVNDQICKYSIPPDVEBKRLLEMPYOEQLLREHYOKKFNKSTYSRSVDVLYTFANC 180
QY	181	GLDIFGIALLRTRADLQWNSNAQLLDYCSKGYNISWELQNEPNSFLKKADIFINGS 240

DB 181 GIDLIFGIALALRTADLQWSSNAQLLDYSSKGNINISWEIGNEPNSFLKKAADIFINGS 240
 QY 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVQOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300
 DB 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVQOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300
 QY 301 NGRTATREDPLNPVDLDFISSVQKVFQVVESTRPKKVLGSETSSAYGGAPLLSDTFA 360
 DB 301 NGRTATREDPLNPVDLDFISSVQKVFQVVESTRPKKVLGSETSSAYGGAPLLSDTFA 360
 QY 361 AGFWMLDKLGISAPMGIEVVMQVFFGAGNHYLVNDENPDLPDWLSILFFKLVGKTYLM 420
 DB 361 AGFWMLDKLGISAPMGIEVVMQVFFGAGNHYLVNDENPDLPDWLSILFFKLVGKTYLM 420
 QY 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAIVLNHVTYRLPYFPFSNKQVDKTYL 480
 DB 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAIVLNHVTYRLPYFPFSNKQVDKTYL 480
 QY 481 RPLGPHGLSKSVQVQNGTLTKAVDDOTLPPLMEKPLRGSSSLGPAFYSFFVIRNAKYA 540
 DB 481 RPLGPHGLSKSVQVQNGTLTKAVDDOTLPPLMEKPLRGSSSLGPAFYSFFVIRNAKYA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 3
 AAY57590 standard; Protein; 543 AA.

AC AAY57590;
 DT 02-MAR-2000 (first entry)
 DE Human heparanase.

Human heparanase; hpa; genetic modification; expression; anticancer; angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumor; anti-atherosclerotic; anti-inflammatory; antineurodegeneration; heparin sulphate; heparin-binding growth factor; tumour angiogenesis; metastasis; wound healing; restenosis; atherosclerosis; inflammation; neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis; micrometastasis; autoimmune lesion; kidney failure.

OS Homo sapiens.
 XX W09557244-A1.
 XX PD 11-NOV-1999.
 XX PE 29-APR-1999; 99WC-US09256.
 XX PR 01-MAY-1998; 98US-0071618.
 XX PR 02-MAR-1999; 99US-0260038.
 XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX PA (FRIB-) FRIEDMAN M M.
 XX PI Ben-Artzi H, Ayal-HersHKovitz M, Yacoby-Zeevi O, Pecker I, Peleg Y, Shlom Y;
 DR MPI: 2000-062144/05.
 DR N-PSDB; AAC39195.
 XX Engineered cells that express recombinant heparanase, useful for therapeutically, e.g. for treating angiogenesis and to screen for specific inhibitors, potential anticancer agents
 PS Claim 3; Page 107-109; 118pp; English.
 CC The present invention describes genetically modified cells (A) containing

CC a polynucleotide (I) that encodes a polypeptide with heparanase activity, and expresses recombinant heparanase (II). Heparanase cleaves heparan sulphate (HS) at specific intrachain sites, resulting in release of heparin-binding growth factors, enzymes and proteins that are sequestered by HS in basement membranes, extracellular matrix or cell surfaces. It may also be implicated in tumour angiogenesis and metastases. (II) is potentially useful in wound healing and for treating angiogenesis, restenosis, atherosclerosis, inflammation, neurodegeneration, viral infection and cystic fibrosis. It can also be used to neutralise heparin (an alternative to procaine) and to screen for specific inhibitors (potentially useful for treating cancer and metastases). Antibodies raised against (II) are used for immunodetection and diagnosis of micrometastases, autoimmune lesions and kidney failure. (A) provide (II) in large quantities, in a form that is homogeneously processed and activated/neutralised by a dedicated protease. The present sequence represents human heparanase.

CC Sequence 543 AA;
 SQ

Query Match 100.0%; Score 2842; DB 21; Length 543;
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPLPPPLMLLGLGPIGSLPGALPRPAQADVDLDFPFOEPLHLVSPSFLSVT 60
 DB 1 MLRSKPLPPPLMLLGLGPIGSLPGALPRPAQADVDLDFPFOEPLHLVSPSFLSVT 60
 QY 61 IDANLATDPRFLILGSPKRLTLAGLSPAYLRFSGTIDFLFPKKESTFEERSYQOS 120
 DB 61 IDANLATDPRFLILGSPKRLTLAGLSPAYLRFSGTIDFLFPKKESTFEERSYQOS 120
 QY 121 QVNODICKYGSIPDVEBKRLLEWYQEQLLREHYQKKFKNSTYSRSSVDVLYTFANCS 180
 DB 121 QVNODICKYGSIPDVEBKRLLEWYQEQLLREHYQKKFKNSTYSRSSVDVLYTFANCS 180
 QY 121 QVNODICKYGSIPDVEBKRLLEWYQEQLLREHYQKKFKNSTYSRSSVDVLYTFANCS 180
 DB 121 QVNODICKYGSIPDVEBKRLLEWYQEQLLREHYQKKFKNSTYSRSSVDVLYTFANCS 180
 QY 181 GIDLIFGIALALRTADLQWSSNAQLLDYSSKGNINISWEIGNEPNSFLKKAADIFINGS 240
 DB 181 GIDLIFGIALALRTADLQWSSNAQLLDYSSKGNINISWEIGNEPNSFLKKAADIFINGS 240
 QY 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVQOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300
 DB 241 QLEGEDYIOLHLKLRKSTFKNAKLYGPDVQOPRRKTAAMKLSFLKAGGEVIDSVTHHHYLL 300
 QY 301 NGRTATREDPLNPVDLDFISSVQKVFQVVESTRPKKVLGSETSSAYGGAPLLSDTFA 360
 DB 301 NGRTATREDPLNPVDLDFISSVQKVFQVVESTRPKKVLGSETSSAYGGAPLLSDTFA 360
 QY 361 AGFWMLDKLGISAPMGIEVVMQVFFGAGNHYLVNDENPDLPDWLSILFFKLVGKTYLM 420
 DB 361 AGFWMLDKLGISAPMGIEVVMQVFFGAGNHYLVNDENPDLPDWLSILFFKLVGKTYLM 420
 QY 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAIVLNHVTYRLPYFPFSNKQVDKTYL 480
 DB 421 ASVQSSKRRKRLRVYLHCTNTDNPVRYEGDLTYAIVLNHVTYRLPYFPFSNKQVDKTYL 480
 QY 481 RPLGPHGLSKSVQVQNGTLTKAVDDOTLPPLMEKPLRGSSSLGPAFYSFFVIRNAKYA 540
 DB 481 RPLGPHGLSKSVQVQNGTLTKAVDDOTLPPLMEKPLRGSSSLGPAFYSFFVIRNAKYA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 4
 AAY52990 standard; Protein; 543 AA.

ID AAY52990;
 DT 21-FEB-2000 (first entry)
 DE Human heparanase protein sequence.

XX Human: heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
 KW anti-infective; immunomodulatory; anti-inflammatory; nephrotoxic;
 KW metacarcinoma; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
 KW autoimmune disease; anticancer; kidney disease.
 XX
 OS Homo sapiens.
 XX
 PN MO9957153-A1.
 XX
 PD 11-NOV-1999.
 XX
 PF 29-APR-1999; 99MO-US09255.
 XX
 PR 01-MAY-1998; 98US-0071739.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX
 PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
 XX
 DR WPI: 2000-052934/04.
 XX
 DR N-PDB; AA233290.
 XX
 PT Heparanase-specific molecular probes useful for diagnosis and
 PT treatment, e.g. of tumours, and for targeted drug delivery
 XX
 PS Example; Page 81-82; 90pp; English.
 XX
 CC The present invention describes heparanase-specific molecular probes,
 CC useful for methods of detecting heparanase in situ. The probes and
 CC anti-heparanase antibodies are used to detect or quantify the expression
 CC of heparanase, for diagnosis and monitoring of diseases (especially
 CC metastasis), for treatment of heparanase-associated diseases (e.g.
 CC tumours, (adenocarcinoma, squamous cell carcinoma, teratocarcinoma,
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
 CC metastases) derived from liver, prostate, bladder, breast, ovary,
 CC cervix, colon, skin, intestine, stomach, uterus and pancreas, kidney
 CC disease, diabetes and inflammation, haemorrhagic nephritis, nephrotic
 CC syndrome, sepsis and inflammatory or autoimmune disease), for targeted
 CC drug delivery (e.g. of anticancer agents) and as research reagents.
 CC The present sequence represents human heparanase, which is used in the
 CC exemplification of the present invention.
 CC
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 2842; DB 21; Length 543;
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLRSKRALPPMLLLGLPLGALPPPAQADVDLDFTOEPLHLVSPSFLSVT 60
 DB 1 MLRSKRALPPMLLLGLPLGALPPPAQADVDLDFTOEPLHLVSPSFLSVT 60
 QY 61 IDANLATDPPFLILGSPKRTIARGLSPAYLRFGTGTDFLIPDKKSTFEERSYWG 120
 DB 61 IDANLATDPPFLILGSPKRTIARGLSPAYLRFGTGTDFLIPDKKSTFEERSYWG 120
 QY 121 QVNVODICKYSSIPDVBEKRLTEMPYQOLLEBHYQKKRNTYSRSSVDVLYTPANC 180
 DB 121 QVNVODICKYSSIPDVBEKRLTEMPYQOLLEBHYQKKRNTYSRSSVDVLYTPANC 180
 QY 181 GDPLIFGINALRTADIQMNSSNAQLLDYCSSKGYNISWELNEPNSFIKKADIFING 240
 DB 181 GDPLIFGINALRTADIQMNSSNAQLLDYCSSKGYNISWELNEPNSFIKKADIFING 240
 QY 241 QLGSDYIQLHLKLRKSTFKNAKLYGPDVGPQRRRTAKMLKSLKAGGEVIDSVTWHYYL 300
 DB 241 QLGSDYIQLHLKLRKSTFKNAKLYGPDVGPQRRRTAKMLKSLKAGGEVIDSVTWHYYL 300

QY 301 NGRTATREDPLNPVLDIFISSVQKFOVVESTRGKKVWLGETSSAYGAGPALLSDTFA 360
 DB 301 NGRTATREDPLNPVLDIFISSVQKFOVVESTRGKKVWLGETSSAYGAGPALLSDTFA 360
 QY 361 AGFWMLDLKGLSARWGLEVNROVFGAGNYHLVDENDDPLPDYWLSTLFFKKLVGTXYLM 420
 DB 361 AGFWMLDLKGLSARWGLEVNROVFGAGNYHLVDENDDPLPDYWLSTLFFKKLVGTXYLM 420
 QY 421 ASVQSKRRKRLRYVILHCTNTNPNRYKGGDLTLVAIINLHNTKYRLRPFSNKVDKXLL 480
 DB 421 ASVQSKRRKRLRYVILHCTNTNPNRYKGGDLTLVAIINLHNTKYRLRPFSNKVDKXLL 480
 QY 481 RPLGPHGLSSKSVQNLGTLTKWVDQTLPLMEKPLRPGSSGLPAPFSYEFFVIRNAVYA 540
 DB 481 RPLGPHGLSSKSVQNLGTLTKWVDQTLPLMEKPLRPGSSGLPAPFSYEFFVIRNAVYA 540
 QY 541 ACT 543
 DB 541 ACT 543
 RSBLT 5
 AA97635
 ID AA97635 standard; Protein; 543 AA.
 XX
 AC AA97635;
 XX
 DT 20-APR-2001 (first entry)
 XX
 DE Human heparanase protein sequence.
 XX
 KW Heparanase; hnppl; wound healing; angiogenesis; restenosis; Scarpa;
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
 KW gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200100643-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 19-JUN-2000; 2000MO-IL00358.
 XX
 PR 25-JUN-1999; 99US-0140801.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA Pecker I, Michal I, Itzhaki H;
 XX
 PI WPI: 2001-137930/14.
 XX
 DR New polynucleotides and polypeptides that are distantly homologous to
 PT heparanase, useful in wound healing, as well as in gene therapy
 PT protocols for angiogenesis, restenosis, atherosclerosis, or
 PT inflammation -
 XX
 PS Disclosure; Page 64-65; 67pp; English.
 XX
 CC This sequence represents a heparanase of the invention.
 CC The heparanase DNA and protein sequences are useful in wound healing,
 CC angiogenesis, restenosis, atherosclerosis, inflammation, pulmonary
 CC disease, neurodegenerative diseases (such as Scarpa, Alzheimer's
 CC disease, and Creutzfeldt-Jakob disease) or viral infections. The
 CC heparanase coding sequence is particularly useful in gene therapy.
 CC
 SO Sequence 543 AA;
 Query Match 100.0%; Score 2842; DB 22; Length 543;
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLRSKRALPPMLLLGLPLGALPPPAQADVDLDFTOEPLHLVSPSFLSVT 60

DB 1 MLRSKRALPPPLMLLLGPISSGALPRPAQADVDVDFPTOEPHLVSPSLST 60
 QY 61 IDANLATDPREFLLILGSPKRLTLARGLSPAYLRFGGTKTDPLIFDKKSTEEERYS 120
 DB 61 IDANLATDPREFLLILGSPKRLTLARGLSPAYLRFGGTKTDPLIFDKKSTEEERYS 120
 QY 121 QVNODICKGSIIPDVEEKLRLMPYQOBLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
 DB 121 QVNODICKGSIIPDVEEKLRLMPYQOBLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
 QY 181 GDLIFGALNALTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240
 DB 181 GDLIFGALNALTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240
 QY 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVQPRKRTAKMLKSPFKAGGEVIDSVTHHYYL 300
 DB 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVQPRKRTAKMLKSPFKAGGEVIDSVTHHYYL 300
 QY 301 NGRTATREDPFLNDVDLIFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360
 DB 301 NGRTATREDPFLNDVDLIFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360
 QY 361 AGFMWLDKLGLSARMGIEVVMROVFFGAGNYHLVDENFDLPYMWLSLFFKLVGKVL 420
 DB 361 AGFMWLDKLGLSARMGIEVVMROVFFGAGNYHLVDENFDLPYMWLSLFFKLVGKVL 420
 QY 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTLVAINLHNTKYLRLPYFSSNKQVDKYL 480
 DB 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTLVAINLHNTKYLRLPYFSSNKQVDKYL 480
 QY 481 RPLGPHGLSKSVQVQNGLTLMVDDOTLPLMEKPLRPSSGLPAPFSYSPFVIRNAKVA 540
 DB 481 RPLGPHGLSKSVQVQNGLTLMVDDOTLPLMEKPLRPSSGLPAPFSYSPFVIRNAKVA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 6
 ABB07813
 ID ABB07813 standard; protein: 543 AA.

AC ABB07813;
 DT 03-JUL-2002 (first entry)
 DE Human heparanase sequence.
 KW Heparanase; catalytic; cytosolic; antiviral; antibacterial; enzyme;
 KM anti-protozoan; neuroprotective; heparin; human.
 OS Homo sapiens.
 XX
 XX Key location/Qualifiers
 FT Peptide 1..33
 FT /note= "signal peptide"
 FT Protein 36..543
 FT /note= "mature protein"
 US2002034810-A1.
 PD 21-MAR-2002.
 PF 16-AUG-2001; 2001US-0930218.
 PR 20-SEP-2000; 2000US-0666390.
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PI Goldsmith O, Pecker I, Vlodevsky I, Michal I, Zcharia E;
 XX

DR WPI: 2002-338926/37
 XX Nucleic acid encoding avian and reptile heparanase polypeptide is
 PT useful to treat various heparin-related disorders and the signal
 PT peptide is useful in production of membrane-targeted or secreted
 PT recombinant proteins
 XX
 PS Disclosure: Fig 1a; 39pp; English.
 XX
 CC The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumor angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a human heparanase protein sequence used in similarity
 CC studies.
 CC
 CC Sequence 543 AA;
 SQ
 Query Match 100.0%; Score 2842; DB 23; Length 543;
 Best Local Similarity 100.0%; Pred. No. 9.5e-274;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKRALPPPLMLLLGPISSGALPRPAQADVDVDFPTOEPHLVSPSLST 60
 DB 1 MLRSKRALPPPLMLLLGPISSGALPRPAQADVDVDFPTOEPHLVSPSLST 60
 QY 61 IDANLATDPREFLLILGSPKRLTLARGLSPAYLRFGGTKTDPLIFDKKSTEEERYS 120
 DB 61 IDANLATDPREFLLILGSPKRLTLARGLSPAYLRFGGTKTDPLIFDKKSTEEERYS 120
 QY 121 QVNODICKGSIIPDVEEKLRLMPYQOBLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
 DB 121 QVNODICKGSIIPDVEEKLRLMPYQOBLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
 QY 181 GDLIFGALNALTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240
 DB 181 GDLIFGALNALTADLQWSSNAQLLDYCSKGYNISWEIGNEPNSFLKKADIFINGS 240
 QY 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVQPRKRTAKMLKSPFKAGGEVIDSVTHHYYL 300
 DB 241 QLGEDYIQLHKLRLKSTFKNAKLYGPDVQPRKRTAKMLKSPFKAGGEVIDSVTHHYYL 300
 QY 301 NGRTATREDPFLNDVDLIFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360
 DB 301 NGRTATREDPFLNDVDLIFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLSDTPA 360
 QY 361 AGFMWLDKLGLSARMGIEVVMROVFFGAGNYHLVDENFDLPYMWLSLFFKLVGKVL 420
 DB 361 AGFMWLDKLGLSARMGIEVVMROVFFGAGNYHLVDENFDLPYMWLSLFFKLVGKVL 420
 QY 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTLVAINLHNTKYLRLPYFSSNKQVDKYL 480
 DB 421 ASVQSKRRRLRYVLAHCTNTDNPYKEGDLTLVAINLHNTKYLRLPYFSSNKQVDKYL 480
 QY 481 RPLGPHGLSKSVQVQNGLTLMVDDOTLPLMEKPLRPSSGLPAPFSYSPFVIRNAKVA 540
 DB 481 RPLGPHGLSKSVQVQNGLTLMVDDOTLPLMEKPLRPSSGLPAPFSYSPFVIRNAKVA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 7
 AAY02346
 ID AAY02346 standard; protein: 592 AA.

XX

AC AAY02346;
 XX
 DT 09-JUL-1999 (first entry)
 XX
 DE A human heparanase protein.
 XX
 KW Heparanase: hpa; modulator; heparin-binding growth factor;
 KW cellular response; cytokine; cell interaction; plasma lipoprotein;
 KW cellular susceptibility; infection; disintegration;
 KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
 KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
 KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.
 XX
 OS Homo-sapiens.
 EN MO9111798-A1.
 XX
 PD 11-MAR-1999.
 XX
 PF 31-AUG-1998; 98MO-US17954.
 XX
 PR 02-JUL-1998; 98US-0109386.
 PR 02-SEP-1997; 97US-0922170.
 XX
 XX (FRIE/) FRIEDMAN M M.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (INST-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 PI Feinstein E, Pecker I, Vlodavsky I;
 XX
 DR WPI; 1999-302255/25.
 DR N-PSDB; AAX35650.
 XX
 PT New human polynucleotide useful for treating angiogenesis,
 PT restenosis, and inflammation
 XX
 XX Claim 6; Page 65-66; 63pp; English.
 XX
 CC The specification describes a polypeptide having heparanase (hpa)
 CC activity. The recombinant protein is used as a modulator of
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoal and bacterial infections
 CC or disintegration of neurodegenerative plaques. Heparanase may be
 CC useful for conditions such as wound healing, angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
 CC infections. Mammalian heparanase can be used to neutralize plasma
 CC heparin, and anti-heparanase antibodies may be applied for
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
 CC and renal failure in biopsy specimens, plasma samples, and body fluids.
 CC The present sequence represents human heparanase.
 XX
 XX Sequence 592 AA;
 XX
 Query Match 100.0%; Score 2842; DB 20; Length 592;
 Best Local Similarity 100.0%; Pred. No. 1,1e-273; Indels 0; Gaps 0;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLRSKRLPPLMLLGLPLSPALPRPAQCVVDLDFPTQPHLWSPFLSVT 60
 DB 50 MLRSKRLPPLMLLGLPLSPALPRPAQCVVDLDFPTQPHLWSPFLSVT 109
 QY 61 IDANLATDPFLILGSPKRTLRGLSPVYLRFGTKTDFLIDPKCESTFEERSYWG 120
 DB 110 IDANLATDPFLILGSPKRTLRGLSPVYLRFGTKTDFLIDPKCESTFEERSYWG 169
 QY 121 QNVADICKYSIPDVREKRLLEMPYOBLLREHYOKKFKNSYSSVDVLYTFPANC 180
 DB 170 QNVADICKYSIPDVREKRLLEMPYOBLLREHYOKKFKNSYSSVDVLYTFPANC 229
 QY 181 GLDLIFGLNLRLTRADLQNNSSNAQLLDVCCSSKGYNISWELGNEPNSFLKADIFPNC 240
 DB 230 GLDLIFGLNLRLTRADLQNNSSNAQLLDVCCSSKGYNISWELGNEPNSFLKADIFPNC 289

QY 241 QUGEDYIQLHLKLRKSTFKNAKLYGPDVGOPRRRTAKKLSFLKAGCEVIDSVTHHYLL 300
 DB 290 QUGEDYIQLHLKLRKSTFKNAKLYGPDVGOPRRRTAKKLSFLKAGCEVIDSVTHHYLL 349
 QY 301 NGRTATREDPLNPDLIDFISSVQKVFQVVESTPRGKVMIGETSSAYGGAPELLSDTFA 360
 DB 350 NGRTATREDPLNPDLIDFISSVQKVFQVVESTPRGKVMIGETSSAYGGAPELLSDTFA 409
 QY 361 AGFWMLDKLGLSARWGIEVVRQVFPAGNYHLVDENFDPLDYWLSLLEFKLVGTQKLM 420
 DB 410 AGFWMLDKLGLSARWGIEVVRQVFPAGNYHLVDENFDPLDYWLSLLEFKLVGTQKLM 469
 QY 421 ASVQSKRRKRLRVYHACNTNPNPKYEGDILLVAILNLNNTKYLRLPFPENKQVDKYL 480
 DB 470 ASVQSKRRKRLRVYHACNTNPNPKYEGDILLVAILNLNNTKYLRLPFPENKQVDKYL 529
 QY 481 RPLGPHGLSKSVQVQNGTLTKVVDQTLPLMEKPLRPGSSILGLPAPSYSPFVIRNAKVA 540
 DB 530 RPLGPHGLSKSVQVQNGTLTKVVDQTLPLMEKPLRPGSSILGLPAPSYSPFVIRNAKVA 589
 QY 541 ACT 543
 DB 590 ACT 592
 RESULT 8
 AAB08850
 ID AAB08850 standard; Protein; 592 AA.
 XX
 XX AAB08850;
 AC
 AC 15-JAN-2001 (first entry)
 DT
 XX
 XX Amino acid sequence of a human heparanase polypeptide.
 DE
 XX
 KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
 KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
 KW wound healing; infection; burn; angiogenesis; restenosis;
 KW atherosclerosis; inflammation; neurodegenerative disease;
 KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
 XX
 OS Homo sapiens.
 XX
 PN MO20005178-A1.
 PD 08-SEP-2000.
 XX
 PF 14-FEB-2000; 2000MO-US03542.
 XX
 PR 01-MAR-1999; 99US-0258892.
 XX
 XX (INST-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV
 PA (FRIE/) FRIEDMAN M M.
 XX
 PI Pecker I, Vlodavsky I, Feinstein E;
 XX
 DR WPI; 2000-579289/54.
 DR N-PSDB; AAA75053.
 XX
 PT New polynucleotides encoding a polypeptide having heparanase activity,
 PT useful in wound healing and in gene therapy, particularly in treating
 PT tumour, inflammation, autoimmunity, neurodegenerative diseases
 XX
 XX Claim 22; Page 122-123; 152pp; English.
 XX
 CC The present sequence represents a human protein with heparanase catalytic
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
 CC particularly in treating tumour, inflammation or autoimmunity.
 CC Particularly, the polynucleotide is useful in modulating the
 CC bioavailability of heparin-binding growth factors, cellular responses
 CC to heparin-binding growth factors (e.g. bFGF) and cytokines

QY 481 RPLGPHGLSKSVQVNGLTAKVDDOTLPPLMEKPLRPSSGLGPAFYSFFVIRNAKYA 540
 DB 481 RPLGPHGLSKSVQVNGLTAKVDDOTLPPLMEKPLRPSSGLGPAFYSFFVIRNAKYA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 10

AAB86206 standard; Protein; 543 AA.

AC AAB86206;

DT 24-AUG-2001 (first entry)

DE Human heparanase inhibitor protein.

XX Heparanase; inhibitor; cardiac insufficiency; cardiac; nephrotropic;
 KW hepatocytic; veterinary medicine; congestive heart failure; dyspnoea;
 KW primary cardiomyopathy; peripheral edema; pulmonary congestion;
 KW hepatic congestion; hydrothorax; ascites; nocturia; human.

OS Homo sapiens.

PN DE19955803-A1.

PD 23-MAY-2001.

PF 19-NOV-1999; 99DE-1055803.

PR 19-NOV-1999; 99DE-1055803.

XX (KNOL) KNOLL AG.

PI Herr D, Hahn A, Laux V;

XX WPI; 2001-368371/39.

DR N-PSDB; AAH20940.

PT Treatment or prevention of cardiac insufficiency and dyspnoea, comprises
 PT administration of heparanase inhibitor

PS Disclosure; Page 11-13; 16pp; German.

CC This invention describes a novel heparanase inhibitor which can be used
 CC for the treatment or prevention of cardiac insufficiency and associated
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
 CC the invention has cardiac, nephrotropic and hepatocytic activity. The
 CC products of the invention can be used in human and veterinary medicine,
 CC for the treatment or prevention of congestive heart failure e.g. primary
 CC cardiomyopathy. Associated conditions treated or prevented with the
 CC inhibitor are especially peripheral edemas, pulmonary and hepatic
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
 CC nocturia can also be treated. This sequence represents the human
 CC heparanase protein described in the method of the invention.

XX Sequence 543 AA;

Query Match 99.94; Score 2838; DB:22; Length 543;

Best Local Similarity 99.84; Pred. No. 2.4e-273;
 Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPAAPPPMLLLGLGPGALPRPAOAVDVLDPFTOEPHLVPSFLSYT 60
 DB 1 MLRSKPAAPPPMLLLGLGPGALPRPAOAVDVLDPFTOEPHLVPSFLSYT 60
 QY 61 IDANLATDPRFLILGSPKRLTARGLSPAYLRFSGTKTDFLIDPKKESTFEERSYWS 120
 DB 61 IDANLATDPRFLILGSPKRLTARGLSPAYLRFSGTKTDFLIDPKKESTFEERSYWS 120

QY 121 QVNOGICCKGSIPEVEEKLRLMEPYQQLREHYCKFNSTYSRSRVLYTFANCS 180
 DB 121 QVNOGICCKGSIPEVEEKLRLMEPYQQLREHYCKFNSTYSRSRVLYTFANCS 180
 QY 181 GDLIFGALNALLRTADLQWSSNAQLLDYSSKGYNISWELGNEPNSFLKKADI FINGS 240
 DB 181 GDLIFGALNALLRTADLQWSSNAQLLDYSSKGYNISWELGNEPNSFLKKADI FINGS 240
 QY 241 QGDEYIQLKLLRSTFRNAKLVPDVGOPRRRTAKLKSFLRAGGVISVYMHAYL 300
 DB 241 QGDEYIQLKLLRSTFRNAKLVPDVGOPRRRTAKLKSFLRAGGVISVYMHAYL 300
 QY 301 NGRTATREDPLNPVDLDFISSVQKVPFVSTRGKVMGESSAYAGGAPLLSDTFA 360
 DB 301 NGRTATREDPLNPVDLDFISSVQKVPFVSTRGKVMGESSAYAGGAPLLSDTFA 360
 QY 361 AGFWMLDKLGLSARNGIEVVRQVFFGAGNYHLDENFDPLDYWLSILFKKLVTGYLM 420
 DB 361 AGFWMLDKLGLSARNGIEVVRQVFFGAGNYHLDENFDPLDYWLSILFKKLVTGYLM 420
 QY 421 ASVQSKRKRLRYVHCTNTDNPVKEDLTLYANLNNTKYLRLPYPFNKQVDKYL 480
 DB 421 ASVQSKRKRLRYVHCTNTDNPVKEDLTLYANLNNTKYLRLPYPFNKQVDKYL 480
 QY 481 RPLGPHGLSKSVQVNGLTAKVDDOTLPPLMEKPLRPSSGLGPAFYSFFVIRNAKYA 540
 DB 481 RPLGPHGLSKSVQVNGLTAKVDDOTLPPLMEKPLRPSSGLGPAFYSFFVIRNAKYA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 11

AA30124 standard; Protein; 588 AA.

AC AA30124;

DT 20-MAR-2003 (updated)

DT 14-OCT-1999 (first entry)

DE A human protein with heparanase activity.

XX Human; heparanase; heparan sulfate; trauma; autoimmune disease;
 KW skin disease; cardiovascular disease; nervous system disease;
 KW Alzheimer's disease; cancer; cancer metastasis; angiogenesis;
 KW inflammation; arthritis.

OS Homo sapiens.

PN WO9940207-A1.

XX 12-AUG-1999.

PF 05-FEB-1999; 95WO-BP00777.

PR 09-FEB-1998; 98GB-0002725.

XX (NOVS) NOVARTIS AG.

PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Nakajima M, Toyoshima M;

DR WPI; 1999-494300/41.

XX N-PSDB; AAX86671.

PT New heparanase polypeptide useful for treating autoimmune diseases,
 PT skin diseases, cardiovascular diseases and nervous system diseases
 PT including Alzheimer's disease

PS Claim 3; Page 29-31; 40pp; English.

QY 181 GDLIFGLNALRTADLQNNSSNAQLLDYCSKGNYSMEIGNEBNSFLKKAADIFINS 240
 DB 181 GDLIFGLNALRTADLQNNSSNAQLLDYCSKGNYSMEIGNEBNSFLKKAADIFINS 240
 QY 241 QLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMEKSLFKAGGEVIDSVTHMYL 300
 DB 241 QLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMEKSLFKAGGEVIDSVTHMYL 300
 QY 301 NGRTATREDPLNDVDLDFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDTFA 360
 DB 301 NGRTATREDPLNDVDLDFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDTFA 360
 QY 361 AGPMWLDKLGSLRMGIEVVMROVFRGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKVM 420
 DB 361 AGPMWLDKLGSLRMGIEVVMROVFRGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKVM 420
 QY 421 ASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHNVTKYLRLPYFSNKOVDKYL 480
 DB 421 ASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHNVTKYLRLPYFSNKOVDKYL 480
 QY 481 RLPGPHGLLSKSVQNLGLTLKAVDDQTLPLMEKPLRPSSISGLPFAFSYFFVIRNAKYA 540
 DB 481 RLPGPHGLLSKSVQNLGLTLKAVDDQTLPLMEKPLRPSSISGLPFAFSYFFVIRNAKYA 540
 QY 541 ACT 543
 DB 541 ACT 543

RESULT 13

ABP56822
 ID ABP56822 standard; Protein; 545 AA.

AC ABP56822;
 DT 02-APR-2003 (first entry)
 XX
 DE Human heparanase protein SEQ ID NO:18.
 XX
 KW Human; heparanase; phosphotriphosphate; antilysine oligonucleotide;
 XX
 KW cytosolic; gene therapy; tumour; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO2003004705-A1.
 PD 16-JAN-2003.
 XX
 PP 01-JUL-2002; 2002WO-US20636.
 XX
 PR 05-JUL-2001; 2001US-0899440.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Stein C;
 DR WPI; 2003-201558/19.
 DR N-PSDB; AB222816.
 XX
 PT New oligonucleotide having a sequence complementary to a sequence of
 PT ribonucleic acid encoding a heparanase, useful for preparing a
 PT composition for treating tumor.
 XX
 PS Disclosure; Page 46-47; 48pp; English.

CC The present invention describes an oligonucleotide having a sequence
 CC complementary to a sequence of ribonucleic acid encoding a heparanase.
 CC The oligonucleotide hybridises with the ribonucleic acid under conditions
 CC of high stringency and has a sequence comprising 10-40 bp. The
 CC internucleotide linkages of the oligonucleotide comprise at least one
 CC phosphorothioate linkage. Hybridisation of the oligonucleotide to the
 CC ribonucleic acid inhibits expression of the heparanase, where inhibition

CC of heparanase means at least a 50% reduction in the quality of
 CC heparanase. Also described: (1) a method of inhibiting expression of a
 CC heparanase in a cell; (2) a composition comprising the above
 CC oligonucleotide in an amount effective to inhibit the expression of
 CC heparanase in the cell and a carrier; and (3) a method of treating a
 CC tumour in a subject comprising administering to the subject an amount of
 CC the above oligonucleotide effective to inhibit expression of a heparanase
 CC in the subject. Heparanase antisense oligonucleotides have cytostatic
 CC activity, can be used in gene therapy, and can be used for preparing a
 CC composition for treating tumours. The present sequence represents human
 CC heparanase, which is given in the exemplification of the present
 CC invention.

CC Sequence 545 AA;

Query Match 99.1%; Score 2817; DB 24; Length 545;
 Best Local Similarity 99.4%; Pred. No. 3e-771;
 Matches 542; Conservative 1; Mismatches 0; Indels 2; Gaps 2;

QY 1 MLNLSKRALPP-IMLLILGPGLSGALPRPAQA-ODVVDLDFPQEPHLVSPFLS 58
 DB 1 MLNLSKRALPP-IMLLILGPGLSGALPRPAQAODVVDLDFPQEPHLVSPFLS 60
 QY 59 VTIDANLATDRPFLILGSPKLTARGLSPAYLRFPGTKTDFLIPDKKSTEEBSYV 118
 DB 61 VTIDANLATDRPFLILGSPKLTARGLSPAYLRFPGTKTDFLIPDKKSTEEBSYV 120
 QY 119 OSQVNOIDICKYGIIPDVEEKLRLWPYQEQQLLREHYOKKFNKSTYSRSSVDVLYTFAN 178
 DB 121 OSQVNOIDICKYGIIPDVEEKLRLWPYQEQQLLREHYOKKFNKSTYSRSSVDVLYTFAN 180
 QY 179 CSGLDIFGLNALRTADLQNNSSNAQLLDYCSKGNYSMEIGNEBNSFLKKAADIFIN 238
 DB 181 CSGLDIFGLNALRTADLQNNSSNAQLLDYCSKGNYSMEIGNEBNSFLKKAADIFIN 240
 QY 239 GSQLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMEKSLFKAGGEVIDSVTHMY 298
 DB 241 GSQLGEDYIQLHLKLRKSTFKNAKLYGPDVGPARRKTAAMEKSLFKAGGEVIDSVTHMY 300
 QY 299 YLNGRTATREDPLNDVDLDFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDT 358
 DB 301 YLNGRTATREDPLNDVDLDFISSVQKVFQVESTRPCKKWLGETSSAYGGAPLSDT 360
 QY 359 FAAGFMWLDKLGSLRMGIEVVMROVFRGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKV 418
 DB 361 FAAGFMWLDKLGSLRMGIEVVMROVFRGAGNYHLVDENPDLDPYWLSTLFLKLVGKTKV 420
 QY 419 LMASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHNVTKYLRLPYFSNKOVDKY 478
 DB 421 LMASVQSKRRKRLRYLHCTNTDNPARYEGDLTLVAINLHNVTKYLRLPYFSNKOVDKY 480
 QY 479 LLRPLGPHGLLSKSVQNLGLTLKAVDDQTLPLMEKPLRPSSISGLPFAFSYFFVIRNAK 538
 DB 481 LLRPLGPHGLLSKSVQNLGLTLKAVDDQTLPLMEKPLRPSSISGLPFAFSYFFVIRNAK 540
 QY 539 VAACT 543
 DB 541 VAACT 545

RESULT 14

AAV34173
 ID AAV34173 standard; Protein; 530 AA.

AC AAV34173;
 DT 15-NOV-1999 (first entry)
 XX
 DE Human pre-proheparanase protein sequence.
 XX
 KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
 KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;

inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
 tumor growth; fibroproliferative disorder; neurodegenerative disease;
 therapy.

Homo sapiens.
 WO9943830-A2.

02-SEP-1999.
 18-FEB-1999; 99WO-US01489.

26-MAR-1998; 98US-0079401.
 24-FEB-1998; 98US-0075706.

(PHAA) PHARMACIA & UPJOHN CO.

Faibanks MB, Heinrikson RL, Mildner AM,
 WPI; 1999-540598/45.
 N-PSDB; AA211236.

New isolated platelet heparanase polypeptides, used to develop
 products for, e.g. wound healing and blocking angiogenesis

Claim 12; Fig 7; 57pp; English.

This sequence is the human pre-proheparanase of the invention. This
 sequence was isolated from human platelets. The heparanase can be used
 for identifying agents which alter heparanase activity. The heparanase
 can be used for wound healing or for blocking angiogenesis or
 inflammation. It can be used for treating e.g. psoriasis, diabetic
 retinopathy or solid tumors, or for the degradation of heparin and the
 neutralisation of heparin's anticoagulant properties during surgery.
 Inhibitors of heparanase activity can be used in the treatment of
 arthritis, asthma, and other inflammatory diseases, vascular restenosis,
 atherosclerosis, tumor growth and progression, fibroproliferative
 disorders, and central nervous system (CNS) and neurodegenerative
 diseases. The products can also be used for detection and diagnosis. The
 purified heparanase, both recombinantly produced human heparanase and
 heparanase isolated from human platelet activity, allows for the
 convenient selection of compounds having anti-heparanase activity,
 i.e. inhibitors of heparanase activity, by measuring inhibition of
 heparanase activity. Inhibition of heparanase activity can be measured by
 blocking heparanase-mediated release of radioactive fragments from in
 vivo radiolabelled (HSPG)/heparin.

Sequence 530 AA;
 Query Match 97.3%; Score 2764; DB 20; Length 530;
 Best Local Similarity 99.4%; Pred. No. 5.4e-266;
 Matches 527; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

14 MLLGLPGLPSPGALPPPAQADVDVLDLFTFOEPHLAVSPFLSTYTDANATDPRPI 73
 1 MLLGLPGLPSPGALPPPAQADVDVLDLFTFOEPHLAVSPFLSTYTDANATDPRPI 60
 74 LGSPLRTLAGLSAYLRPGSTKDFLIPPKKSTEBBSWMOSONODICKKGSIF 133
 61 LGSPLRTLAGLSAYLRPGSTKDFLIPPKKSTEBBSWMOSONODICKKGSIF 120
 134 PVEBKRLLEWPOEOLREHYOKKFKKSTYSRSSVDLYFANCSGLDLFGNALMR 193
 121 PVEBKRLLEWPOEOLREHYOKKFKKSTYSRSSVDLYFANCSGLDLFGNALMR 180
 194 TADLQNNSSNAQLLDYSSKGNISWEIAGNBNSTLKAADIFINGSQIGBYIQHLK 253
 181 TADLQNNSSNAQLLDYSSKGNISWEIAGNBNSTLKAADIFINGSQIGBYIQHLK 240
 254 RSTFPAALYGDVQGPRTKATKMLKSLKAGGEYIDVTMHHYTLNRTTRBDFLNP 313
 241 RSTFPAALYGDVQGPRTKATKMLKSLKAGGEYIDVTMHHYTLNRTTRBDFLNP 300

314 DVLDIFISSVOKVQVYVESTRPGKWLIGETSSAGGAPLSDTEPAGMMLDKLGLS
 301 DVLDIFISSVOKVQVYVESTRPGKWLIGETSSAGGAPLSDTEPAGMMLDKLGLS 360
 374 RMGIEVMRQVFPAGVYHVDENFDLPDYMILSLFKKLVTGKVLMAVSQSKRRRLRY 433
 361 RMGIEVMRQVFPAGVYHVDENFDLPDYMILSLFKKLVTGKVLMAVSQSKRRRLRY 420
 434 YLHCTNTDNPYKGGDLTLVAINLHNTKYLRLPYPSNKQVDKYLRLPGHGLLSKV 493
 421 YLHCTNTDNPYKGGDLTLVAINLHNTKYLRLPYPSNKQVDKYLRLPGHGLLSKV 480
 494 QLNGLTLKMDVDTPLPLMEKPLRPGSSGLPAPSVSEFPIRNKXVACI 543
 481 QLNGLTLKMDVDTPLPLMEKPLRPGSSGLPAPSVSEFPIRNKXVACI 530

RESULT 15
 AA17083
 ID AA17083 standard; Protein; 532 AA.
 AC AA17083;
 XX 21-JUL-1999 (first entry)
 DT Seq ID No: 15 of WO9921975.
 DE Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
 KM metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
 KM atherosclerosis; atherosclerosis; inflammation; tissue development;
 KM human; HSPG.

OS Homo sapiens.
 PN WO9921975-A1.
 PD 06-MAY-1999.
 PF 28-OCT-1998; 98WO-AU00898.
 XX 09-DEC-1997; 97AU-0000812.
 PR 28-OCT-1997; 97AU-0000062.
 XX (AUS) UNIV AUSTRALIAN NAT.
 PA Freeman CG, Hamdorf BJ, Hulett MD, Parish CR;
 PI WPI; 1999-312956/26.
 DR N-PSDB; AAX37260.
 XX Polynucleotides encoding mammalian endoglucuronidases, especially
 heparanases, useful to promote wound healing

Claim 6; Page 76-79; 112pp; English.

The invention relates to nucleic acid sequences that encode heparanase
 enzymes having endoglucuronidase activity. Recombinant heparanases are
 capable of removing the HS side chain from heparan sulfate proteoglycan
 (HSPG). Sulfated oligosaccharides, sulfonates or HSPG can be used to
 inhibit heparanase, this is useful for treatment of a physiological or
 medical condition associated with elevated heparanase activity, such as
 metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
 atherosclerosis, atherosclerosis and inflammation. The human, murine and
 rat heparanases can be used to enhance wound healing, especially
 associated with tissue development and repair. The conditions mentioned
 above can be diagnosed using specific antibodies, and also using primers
 and probes specific for the heparanase polynucleotides. Other uses of the
 heparanases include sequencing sulfated molecules such as HSPG.

Sequence 532 AA;
 Query Match 96.3%; Score 2737; DB 20; Length 532;
 Best Local Similarity 99.8%; Pred. No. 2.7e-263;

Matches 522; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLLRSPALPPPLMLLLGLPLSPGALPPPAQADVVLDLDFTOEPHLVSPSLSYT	60
DB	1	MLLRSPALPPPLMLLLGLPLSPGALPPPAQADVVLDLDFTOEPHLVSPSLSYT	60
QY	61	IPANLATDRPFLILGSPKLTARGLSPAYLRFGGKTDFLIIPDPKKESTFEERSYQS	120
DB	61	IPANLATDRPFLILGSPKLTARGLSPAYLRFGGKTDFLIIPDPKKESTFEERSYQS	120
QY	121	QVNOODICKYGIIPDVEEKLRLWPYQEOQLLREHYQKKEFNSTYSRSSVDLYTFANCS	180
DB	121	QVNOODICKYGIIPDVEEKLRLWPYQEOQLLREHYQKKEFNSTYSRSSVDLYTFANCS	180
QY	181	GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEBNSFLKADIFINGS	240
DB	181	GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWEIGNEBNSFLKADIFINGS	240
QY	241	QLEGDYIOLHKLKSTFGNAKLYGPDVGOPRRKTAQMLKSFLLKAGGEVIDSVTHHYYL	300
DB	241	QLEGDYIOLHKLKSTFGNAKLYGPDVGOPRRKTAQMLKSFLLKAGGEVIDSVTHHYYL	300
QY	301	NGRTATREDFLPNDVLDIFISSYQKVFQYVESTRPGKVMGETSSAYGGAPLLSDTFA	360
DB	301	NGRTATREDFLPNDVLDIFISSYQKVFQYVESTRPGKVMGETSSAYGGAPLLSDTFA	360
QY	361	AGFMWLDKGLSARNGIEVVMRQVFFGAGNYHLVDENPDPDYMLSLFKLVGTXYLM	420
DB	361	AGFMWLDKGLSARNGIEVVMRQVFFGAGNYHLVDENPDPDYMLSLFKLVGTXYLM	420
QY	421	ASVQSKRRKRLRYLHCTNDNRYKEGDLTYAIVLHNTKYLRPYPEFSNKQYDKYLL	480
DB	421	ASVQSKRRKRLRYLHCTNDNRYKEGDLTYAIVLHNTKYLRPYPEFSNKQYDKYLL	480
QY	481	RPLGPHGLSKSVQVNLGTLKAVDDQTLPLMEKPLRPGSSLG	523
DB	481	RPLGPHGLSKSVQVNLGTLKAVDDQTLPLMEKPLRPGSSLG	523

Search completed: October 22, 2003, 20:24:22
Job time : 88 secs

OM protein - protein search, using SW model

Run on: October 22, 2003, 20:22:55 ; Search time 42 Seconds

(without alignment)

Title: US-09-759-207-2

Sequence: 1 MLRSKPALPPPLMLLLGP.....LPASFSFFVINAVAACT 543

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	897.5	31.6	480 2 JC7506	heparanase protein
2	416	14.6	521 2 T45608	hypothetical prote
3	169.5	6.0	190 2 T01953	hypothetical prote
4	112.5	4.0	356 2 F64383	hypothetical prote
5	111.5	3.9	575 2 T12094	beta-fructofuranos
6	111	3.9	670 2 T10666	hypothetical prote
7	111	3.9	888 2 S32961	hypothetical prote
8	111	3.9	2298 2 T49648	hypothetical prote
9	109.5	3.9	879 2 E91031	probable outer mem
10	108.5	3.8	411 2 S74760	hypothetical prote
11	107.5	3.8	500 2 D87541	beta-xylosidase (1
12	106	3.7	879 2 F85875	probable fibrillar
13	105	3.7	670 2 T38446	microtubule-associ
14	104.5	3.7	788 1 S00652	phosphoribosylamin
15	104	3.7	432 2 F70411	adenylosuccinate s
16	104	3.7	2013 2 A11489	probable peptidogl
17	103.5	3.6	587 2 S36231	beta-fructofuranos
18	103.5	3.6	676 2 AF1153	transcription anti
19	103.5	3.6	687 2 F85188	retrotransposon 11
20	103	3.6	796 2 D97065	transketolase (imp
21	101	3.6	594 2 A82913	hypothetical prote
22	101	3.6	644 2 A97268	methionyl-tRNA syn
23	100.5	3.5	805 2 C86525	DNA gyrase subunit
24	100.5	3.5	805 2 H72098	DNA gyrase, chain
25	100.5	3.5	989 2 AB2140	toxin secretion AB
26	99.5	3.5	510 2 H69893	conserved hypothec
27	99.5	3.5	837 1 A31842	endo-1,4-beta-xyla
28	99	3.5	897 2 G02529	dynein heavy chain
29	99	3.5	4644 1 A38905	dynein heavy chain

30	98.5	3.5	596 2 T04506	hypothetical prote
31	98.5	3.5	629 2 C64180	hypothetical prote
32	98.5	3.5	654 2 T14202	NADH dehydrogenas
33	98.5	3.5	699 2 P59146	DNA topoisomerase
34	98.5	3.5	701 2 D98014	DNA topoisomerase
35	98.5	3.5	746 2 T46821	siderophore recept
36	98.5	3.5	746 2 A05420	Rhizobactin r
37	98.5	3.5	1012 2 UC5925	membrane K10cho pr
38	98	3.4	465 2 T19113	hypothetical prote
39	98	3.4	716 1 C60008	RNA-directed RNA p
40	98	3.4	760 2 T34414	hypothetical prote
41	98	3.4	817 2 H75035	probable membrane
42	97.5	3.4	454 2 T20829	probable serine ca
43	97.5	3.4	511 2 S61166	probable membrane
44	97.5	3.4	604 2 E75119	hypothetical prote
45	97.5	3.4	804 2 G71546	probable DNA gyrase

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human

C/Species: Homo sapiens (man)

C/Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 01-Dec-2000

C/Accession: JC7506

R/McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hancock, M.; Pe

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A/Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family m

A/Reference number: JC7506

A/Accession: JC7506

A/Molecule type: mRNA

A/Residues: 1-480 <MCR>

A/Cross-references: GB:AF282885

C/Comment: This protein, a intracellular membrane-bound enzyme, has biological and ther

C/Genetics:

A/Gene: hpa2a

A/Map position: 10q23-10q24

C/Keywords: heparin binding; membrane bound

Query Match	Best Local Similarity	Score	DB 2	Length	Matches	Conservative	Indels	Gaps
20	PLGPIPGAL	31.6%	897.5	480	202	74	146	139
18	PPACLPALTYLALLLSLSSQAGDRPLVDRAAGLKEKTLILLDSTKNPRTVBN	14.6%	416	521	18	7	13	7
56	FLSVITIDANLADPRFLLGSPKRLTARGLSAPAYRFGSTKTDLLF	6.0%	169.5	190	56	18	13	11
78	FLSLQDLDSIIHD-GMDLFLSKRLVTLARGSPFAFGKRTDFLFQULRNPAKR	4.0%	112.5	356	78	13	13	13
112	FEERSYOSQVNDICCKGSLPPDVEEKLLEWYQEBLLREHYQKFKNSTYSRSSVD	3.9%	111.5	575	112	13	13	13
136	-----GCGPDP-----	3.9%	111	670	136	13	13	13
172	VLYTFANGSGDLIFGNALARTADLQWSSNAQLLDYCSSKGVNIMWEGNEPNSFLK	3.9%	111	2298	172	13	13	13
149	-----DEPNRT	3.9%	111	879	149	13	13	13
232	KADIFINGSQLEBDYIQHLKLR-STFKNAKLYGPDVGQPRRTAKMLKAGLGEVI	3.9%	111	879	232	13	13	13
157	MGRVAVNSQKQKYOIKSLQPIRISRASLIGPNIGRRKRVIALLDGPMKAGSTV	3.9%	111	2298	157	13	13	13
291	DSVTHHTYLLNGRATREDPLNPVDLFISSVQKVPQVESTPBGKVMGRTSSAYG	3.9%	111	879	291	13	13	13
217	DAVWQHCYIDGRVVKWDFKTRLLDPLSDQIRKIQCVVTVYFGKKIMEGVVTSAG	3.9%	111	670	217	13	13	13
351	GAPLLSDPFAAGFWMLDLGASARGIEVVMKROVFFGAGNHLVDENPDPLDYSLIF	3.9%	111	2298	351	13	13	13
277	GTNNLSDSYAGFIMLNTLGNLANOGIDIVIRHSFFDHGYNHLVDQNPPLDPTWLSLY	3.9%	111	879	277	13	13	13

QY 411 KLVGTVKVLMAASVQSGSKRR-----KLRVYVHCTNTDNPYKGGDLTLVAINTANVT 461
 DB 337 KLIGKVLAVNVAAGQKRRPRGRVTRDRLRIYAHCTNNHNYVAGSITLFLINLRSR 396
 QY 462 KTLRLPYPSNQVDYKLLRPUPHGLSKSVQVNLTKRMVDQTLPLMEKPLRPSS 521
 DB 397 KIKLGLAGTRDLKLVHGYLLQPYGOEELKSKSVQVNLQPLVMVDGTLPELKKRPLRAGRT 456
 QY 522 LGLPAPSYSPFYIRNAKVAAC 542
 DB 457 LVIPVYMGFVYKVNALAC 477
 RESULT 2
 T45608
 Hypothetical protein F13G24.30 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T45608
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voel, M.; Robben, J.; Volckaert, G.; Be
 submitted to the Protein Sequence Database, December 1999
 A:Reference number: Z23009
 A:Accession: T45608
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-521 <BBV>
 A:Cross-references: EMBL:AL133421
 A:Experimental source: cultivar Columbia; BAC clone F13G24
 A:Map position: 5
 A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
 A>Note: F13G24.30
 Query Match 14.6%; Score 416; DB 2; Length 521;
 Best Local Similarity 29.2%; Pred. No. 3,5e-23;
 Matches 154; Conservative 68; Mismatches 184; Indels 122; Gaps 24;
 QY 75 LGSPLRLTARGLSPAYLRFSGTKTDPLFDPKKESTFEERSYQWQVNDLCKXGSIIP 134
 DB 55 LTRPLTLTAIKAPKPLRIRIGSLQDQVIVDGNLKT-----PCR----- 94
 QY 135 DVEBKRLRLEMPYOEQLLRHYOKFKNS---TYSRSV---DLYTFANCSGDLIF 186
 DB 95 -----PQK-----NSGLRFGSKGCLHMRMDELNPLTATGAVTF 132
 QY 187 GLNALLRTADLQ-----WSSNAQLLDYSSSKGYNI-SNELGNEPSEFKKADIFIN 238
 DB 133 GINALLRGRHKLKAGKMGAMDHINQDPLNTYTSKGYIDSWEFGNELSG--SGVCAVS 190
 QY 239 GSQLGEDVYQLHKLKRSKTFKNAKLYGPRVGP-----RKTKAKLKSFLKAGEVISV 293
 DB 191 AELYGDLVLADQVINK-VYKSMWLKPLVAPGGFYEQWYTKLEI---SGPSVADV 246
 QY 294 TWHHYLYNGRT--ATREDFLNPDLVDFISSYQKVF---QVVESTPGKAWLGETSSA 347
 DB 247 THHITNLGSGNPALVYKIMDS-----YISQSKTFKQVNOIIOHGRPASWGESGA 302
 QY 348 YCGAAPLSDTTPAAGFMWLKGLSLAKMGI EYVMQVFFGAGNYHVLDE-NFDPPLDYWL 406
 DB 303 YNSGRHRVSDTFIDSEFWYLDQGLMSARHNTKYCQTLWG-GFYGLEKGTVPMPDDYS 361
 QY 407 SLTFKKLVOTKYLMAASVQSGSKRKLRVYHCTNTDNPYKGGDLTLVAINTANVT 464
 DB 362 ALMLRMLMGKGLAVQTDGP--QLRVVAHCK-----GRAGVTLILLINLSQDPTVS 413
 QY 465 -----RLPYPS--NKQVDKYLRLP--LGPHG--LTSKSVOL 495
 DB 444 VNGINVLVMASSRKKSLDLTKRPFWSIGSKADGYLNRREHYLTPENGVLARKTMYL 473
 QY 496 NGTLTKMDDQTLPLMEKPLRP-GSSGLPAPSYSPFYIRNAKVAAC 542
 DB 474 NKSLSKPLTATGDIPIST-EPVLRVSVNPLNVLPLSMFIVLPNFDASAC 520

RESULT 3
 T01953
 Hypothetical protein T2L5.6 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jan-2000
 C:Accession: T01953
 R:Geisels, C.; Smith, A.; Le, T.
 submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of A. thaliana T2L5.
 A:Reference number: Z14470
 A:Accession: T01953
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-190 <GBL>
 A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392
 A:Experimental source: cultivar Columbia
 A:Genetics:
 A:Map position: 4
 A:Introns: 36/2; 69/3
 A>Note: T2L5.6
 C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6
 Query Match 6.0%; Score 169.5; DB 2; Length 190;
 Best Local Similarity 27.8%; Pred. No. 2.1e-05;
 Matches 54; Conservative 34; Mismatches 57; Indels 49; Gaps 9;
 QY 382 ROVEFGAGNYHVLVD-ENFDPPLDYWLSLFKKLVGTQVLMASVQSGSKRRKLVYLCTNT 440
 DB 12 RQSLIG-GNYGLNTVTFETPNPDYXSLIWRQLMGKRALFTTFSGTK--KIRSYTHCA-- 66
 QY 441 DNPYKGGDLTLVAINTANVT-----TKYRLRLPYPSNQVDYKYLRL 483
 DB 67 ---RQSKG-ITVLLMNDLNTTVAAYELNNSFSLRHTKHK-----SYKASSQLFG-- 115
 QY 484 GPHGL-----SKSVQLNGTLTKMVDQTLPLMEKPLRPSSSLGLPAPFS 528
 DB 116 GPNGVIOEYHHLAKQGNLHSGTMLNGLNGLQVNSMGDPLPIPIHNSTEPTIAPYS 175
 QY 529 YSPFYIRNAKVAAC 542
 DB 176 IVFVHRVNVVVPAC 189
 RESULT 4
 F64383
 Hypothetical protein MJO670 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: F64383
 R:Bult, C.V.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
 Reisch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
 Iron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.P.; Roberts, K.M.; Huze, M.A.
 Science 273, 1058-1072, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
 Article: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: F64383
 A:Molecule type: DNA
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Residues: 1-356 <BLU>
 A:Cross-references: GB:U67514; GB:U77117; NID:g2826304; PIDN:AA898664.1; PID:g1591384; 1
 A:Map position: REV596956-595866
 A:Start codon: GTG
 Query Match 4.0%; Score 112.5; DB 2; Length 356;
 Best Local Similarity 21.2%; Pred. No. 1;
 Matches 85; Conservative 48; Mismatches 143; Indels 125; Gaps 18;
 QY 126 ICKT-----GSIPDVEKRLRLEMPYOEQLLRHYOKKFRNSTYSSVD----- 171

Db 14 IRKTKITNGKREKIKERLIKE-----LKEHVLYVETEDGTYTLAAEDDEEMMSKV 66
 172 -----VLYTFANCGLDILFGLMALLRTADLOMSSNMQLLDYSSKGYINISWELAGNP 226
 67 GALKEAIYKFAKPS-----KITDL-----SNPR-VLDLCSGMYAIALMHTNK 109
 227 NS-----FL-----KKADITINGSGLEDYIQLHKLKSTF 258
 110 NAEIDMVEICSEVLEPLFLDIPYKEHEIIMDKVREYFLN-KIGIEY-----KSDY 159
 259 KNAKLYGVDGQPRKTKAMLSPLKAGEVIDSYTMHHYLYNGRTAT--REDFLNPDL 316
 160 DNINLY--VGDARFIIKSDKY-----NVFFDARSPPKDPFLTYTDFL----- 202
 317 DIFSSVQKRVQVVESTPRGKKVWLGRTSSAYGAGAPLLSDTFAAGFMWLDKLGASABG 376
 203 -----KEIKRMBDN--GVLI-----SVSSALIPRSALVDCGFYISEKESVGRKRG 246
 377 IEVMKQVFPFAGNTHLVNDENP-----PLPDVMSLLEPKLVGTVMASVQSGSR 429
 247 ITLAVKKNFKNRINEVDERVIATLALPYRDETLTLTKDKIIEDEERERKLEKLI 306
 430 KLRVYIHCNTNDNPRYKSGDLTLVA--INLHVTKYLRPY 468
 307 KIGTLSTKQIKKGNIPBEILKIQKEDLANSSEIIRKMLKF 347

RESULT 5
 112094
 beta-fructofuranosidase (BC 3.2.1.26) - fava bean
 C/Species: Vicia faba (fava bean)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T12094
 R/Weber, H.; Borstjuk, L.; Heim, U.; Buchner, P.; Mobus, U.
 A/Title: Seed coat-associated invertases of Fava bean control both unloading and storage
 A/Reference number: Z17416; MUID:96093423; PMID:8535137
 A/Accession: T12094
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-575 <WEB>
 A/Cross-references: EMBL:Z35162; NID:g861154; PION:CAA6526.1; PID:g861155
 A/Experimental source: cv. Fido, seed coat
 C/Genetics:
 A/Gene: CWINV1
 C/Superfamily: beta-fructofuranosidase
 C/Keywords: cell wall; glycoprotein; glycosidase; hydrolase

Query Match 3.9%; Score 111.5; DB 2; Length 575;
 Best Local Similarity 21.4%; Pred. No. 2.5;
 Matches 72; Conservative 48; Mismatches 107; Indels 109; Gaps 19;

46 QEPILVLS-----PSFLSVITIDANLTPRFLILLSPKRLTLARGLS-----P 89
 228 KPHISAKRTGMBCPDFPVSLBKGKGLD--LSMGMGNVVKHVKNSLDITREYTYTG 285
 90 AYLR-----FGGKTDF-----LTFPKKSTFEERSYV--OSQVNO 124
 286 TYLQNDKTIIDPKISEDMGGLRYVGNFYASKSFDDPK--NRIIMGANESDKE 341
 125 DICRYG--SIPDV-----EEKLRLEMPYQEQLLR--EHYQKPKNSTYSRSV 170
 342 DDVKVMAGIQAIPIRTWLDSSRQLR-QMPVEELNRLRGQVEKMKRLKKGYG--L 396
 171 DVLTFPANCGLDILFGLMALLRTADLOMSSNMQLLDYSSKGYINISWELAGNP 230
 397 EYKGTASQADVEVFSFSSLDKALFPPNMENAR--DLCAQKSSKVRGGV--PGLL 451
 231 KKADIFINGSGLEDYIQLHKLKSTF-----HKL-----RKSTFKNAKLYGP-----DV 268
 452 TLA-----SKLSEYTSVFFRVFKANHKILMCSDAKSSSLNLEIVKPSFAGFVNDL 505
 269 GQPRKTKAMLSPLKAGEVIDSYTMHHYLYNGRT 304

Db 506 GNNKLSLRSL-----IDHSVSEFGVGCKT 531

RESULT 6
 110666
 hypothetical protein F6E21.40 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Feb-2001
 C/Accession: T10666
 R/Beyan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro
 submitted to the Protein Sequence Database, June 1999
 A/Reference number: Z16533
 A/Accession: T10666
 A/Molecule type: DNA
 A/Residues: 1-670 <BEV>
 A/Cross-references: EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.40
 A/Experimental source: Cultivar Columbia; BAC clone F6E21
 C/Genetics:
 A/Gene: ATSP:F6E21.40
 A/Map position: 4
 A/Intons: 47/3; 87/1; 123/3; 203/3; 230/2; 255/3; 284/3; 305/1; 335/3; 347/3; 370/3; 35
 C/Superfamily: Schizosaccharomyces pombe negative regulator of mitosis skbl

Query Match 3.9%; Score 111; DB 2; Length 670;
 Best Local Similarity 22.4%; Pred. No. 3.4;
 Matches 123; Conservative 77; Mismatches 194; Indels 156; Gaps 33;

51 LVSPFLSVITIDANLTPRFLILSGSPKRLTLARGSPAY--LRFGTKDFLIPPK- 107
 47 LVDSYSPSLVEGN-GVDTVLPVCGSDLV-----LSPGQMSHVVGKISWIDLDSGD 99
 108 -----KSTFEERSVQSOVNODICKYGISPPVEEKLRLMPYOBOLIREHYQKKF 161
 100 EVLMDSETTLKQIAWATHLSIQMCE--PD-----LTRPHYLAGL 139
 162 NSTYSRSV--DVLY-----FRANGS--GLDILFGLMALLRTADLOMSSNQL 207
 140 RVSCRRSSPISDEFELKITEFNQALTFGSSSLFCLNVIASALKMLRVLPLV--SEGDW 196
 208 LDYSSKGYINISWELGN-----EPNSFLKKA--DIFIN-----GSQLEDYIQLHKL 253
 197 DD--TSEGLANDSWELNMSFRLCEHDSKLSVALDVLSTLPSERTSLGWMGSS-VRAIILS 253
 254 RKSTFKNAKLYGPDVGP--RRKTAUKLSFL-KAGEVIDSYTMHHYLYNGRTATREDF 310
 254 TDAFLTNAR-----GYPCLSKRHQKLIAGFFDHAQVVICGKRVHNLQKPLDSSSGTE 307
 311 LNPVLDIFISSVQKRVQVVESTPRGKKVWLGRTSSAYGAGAPLLSDTFAAGFMWLDKLG 370
 308 KNP--LRIVLDYVAVYLRQKMSLSSEDERIBEGYRDFLOAPLOPLMDNLEAQVETFE-- 362
 371 LSAMGIEVMKQVFPFAGNTHLVNDENPDLPDVMSLLEPKLVGTVMASVQSGSR 429
 363 --RDSVYTIQYQ--RAVEKALVDR--VPDEKASEL-----TYLVAVVGAGRGV 406
 421 -ASVQSGKR--RKLRYVYIHCNTNDNPRYKSGDLTLVAINLHVTK-----YLRLEY 468
 407 RASIQAAEEDRKLVKQV--AVEKNRP-----AVVTLNVLKMGMEVDVITILSCDM 455
 469 PFSN--KQVDKYLRLPYGPHGLSKSVQVNLGLTLKMTDDQTLPLM--EKLAPGSSLG 523
 456 RFWVAPRQADITLVSELGSG--DNLSPCLDQAGQFLKP-DGIS 498
 524 LPARSYFV 533
 499 IFS-STYSFI 507

RESULT 7
 S32961
 hypothetical protein YBR259w - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein YBR1727

C.Species: *Saccharomyces cerevisiae*
 C.Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 19-Apr-2002
 C.Accession: S33961; S64140 - sequence revision
 R.Doisnon, F.; Bileau, N.; Crouzet, M.; Aigle, M.
 Yeast 9, 189-199, 1993
 A.Title: The complete sequence of a 19,482 bp segment located on the right arm of chromosome 9
 A.Reference number: S23346; MUID:93220397; PMID:8465606
 A.Accession: S33961
 A.Status: translation not shown
 A.Molecule type: DNA
 A.Residues: 1-688 <DOI>
 A.Cross-references: EMBL:X70529; NID:91907246; PID:CAA49923.1; PID:9296558
 R.Aigle, M.; Bileau, N.C.; Barthe, C.; Bileau, N.; Crouzet, M.; Doisnon, F.
 submitted to the Protein Sequence Database, August 1994
 A.Reference number: S45940
 A.Accession: S46140
 A.Molecule type: DNA
 A.Residues: 1-688 <AIG>
 A.Cross-references: EMBL:X36128; NID:9536684; PID:CAA85222.1; PID:9536685; MIPS:YBR259w
 C.Genetics:
 A.Cross-references: SGD:S0000463
 C.Map position: 2R
 C.Superfamily: *Saccharomyces cerevisiae* hypothetical protein YBR259w

Query Match 3.9%; Score 111; DB 2; Length 688;
 Best Local Similarity 22.5%; Pred. No. 3.6;
 Matches 67; Conservative 45; Mismatches 94; Indels 92; Gaps 16;
 Oy 126 ICKYGIIPDVEEKLRLLEPYOELLREHYOKKFNSTYSRS-----VDVLYT 175
 164 MAEYSMSKWDSDKQQLQCYMEFRMKECLVKEYNFOLQSKSDPLKELIIPMEKIYV 223
 Oy 176 FANCGGLDIPGLNMLRTADLQWSSN-----AQLLD-----YSSSGY----- 216
 224 -NAC--TDFTEBOYRIDAEALMTSKNVFSSISAVLRDLNMFSAFPPYGEALV 280
 Oy 217 -----NISWEIGNEPNSFLKKA--DIF--INGSOLG--BDYVQLHLKLR----- 255
 281 QDPFHRSLKMSNDKVESLIRALIFNDKFPFNKEQVUTKADGIFFLRLAKRKHEHN 340
 Oy 236 -----STFKN--AKLYGPDVGQPRKRTAKMLKSLKAGEV-----IDSV 293
 341 DVKDFHIQVIKYLNSQFKNNYSTLMTSSKTQDRRSHNPPSILDDGNKIGMHWSPIDE- 399
 Oy 234 TMAHYTLG-----RTATREDPLNPVDLDFISSVQVQVVESTR--FGK 338
 400 -YSHFTIDNDEPLMRDKVYKITYTNEQTPPDASALFDS--HKIYALISLRYLPEKR 454
 Db

RESULT 8
 T49648
 Hypothetical protein B8B20.20 [imported] - *Neurospora crassa*
 C.Species: *Neurospora crassa*
 C.Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C.Accession: T49648
 R.Schulze, U.; Aign, V.; Hohensei, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A.Reference number: Z25022
 A.Accession: T49648
 A.Molecule type: DNA
 A.Status: preliminary
 A.Residues: 1-2298 <SCH>
 A.Cross-references: EMBL:AL355933; GSPDB:GN00116; NCSP:B8B20.20
 A.Experimental source: BAC clone B8B20; strain OR74A
 C.Genetics:
 A.Gene: NCSP:B8B20.20
 A.Map position: 6
 A.Introns: 426/3

Query Match 3.9%; Score 111; DB 2; Length 2298;
 Best Local Similarity 19.3%; Pred. No. 22;
 Matches 114; Conservative 79; Mismatches 190; Indels 208; Gaps 28;
 Oy 78 PKRLTLARGLSPA-----YLRFQGTDTDFLIPDPKSTFEERSYQOV-NODIC 127
 1447 PRVDILRLITPSNAHKEACILINRANNOALRVVSGSASRPITRNANVFNQILD 1506
 Oy 128 KYGSIIPDVEEKL-----LEWYQQLREHYOKKFNSTYSRSVDVLYTPAN 178
 1507 QYMAESDIEQFPALGAENNRSIDAAMRELLTKN-----KAVLDLHTSAR 1555
 Oy 179 CSGDLIFGNAU-----LRTDILQ-----MNSNAQLLDYSSKGYNI 218
 1556 AS-LDVAKQAKTEALAYTLNVTOLQKCTTLHFGSGFPMGLINVAL-----DTHAFL 1609
 Oy 219 SW-ELGHEPNSFLKADIFINGSGEDYQLHLKRTFKNAKLYGPDVGQPRRTAK 277
 1610 CMITSSEBOYSSNESSADIDPRQEDAILLLQELKKEPFMA-----RELLAL 1659
 Db 278 MLKSFPLAGSEVID-SVYWHYVINGTATREDPLNDVDLDFISSVQVQVVESTRPG 336
 1660 PLKATTFGKQTEQVACTEKTVTTLAALTAAR-----FLQ--ERYOVLPYEQPG 1706
 Oy 337 K-----KWLGETSSAYGGAFLDTPFAG-----FMWLDK 368
 1707 KYGLPDMKMSGPERRWL-----PLFIATLVNNKVPDEPDITNLSLWQS 1755
 Oy 369 LGLSAR-MGIEVVRQVFPFGAGNYHL--VDENFDPLDPYMLSL-LPKLVG--TKVL-- 419
 1756 IIKRMRFGLTYTIAVLQGRGLPFLAEADVSAQMTDYNHDLFSRAIHYMKALRG 1815
 Db 420 -----MASVQSK-----RRKLRYLHCTNTNPRYKSGDLTLVAINLHNTKYL 464
 1816 ATPAPGVTSASAYGSSASQSIROREBFSH-----PHGLSKSVQNLGLTKMD-- 1847
 Oy 465 RLPPFSNKQVDKYLPLIG-----PHGLSKSVQNLGLTKMD-- 504
 1848 TLQIANTIKQDLPLSLALADPTASSTERHDYMAFTNCLIS-LKSHGVGVVVDSE 1906
 Oy 505 -----DQTLPLMEKPLRPG-----SGLGLPAVSYSF-FVIRAKVA 540
 1907 FLTSDSVSPPLQDPQLHTAGIMAYGVRLSKQDVPASQVLENNPKVA 1957
 Db

RESULT 9
 E91031
 Probable outer membrane protein Ec93221 [imported] - *Escherichia coli* (strain O157:H7, e
 C.Species: *Escherichia coli*
 C.Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C.Accession: E91031
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A.Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
 A.Reference number: A96429; MUID:21156231; PMID:11258796
 A.Accession: E91031
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-879 <HAY>
 A.Cross-references: GB:BA000007; PIDN:BA036644.1; PID:913362691; GSPDB:GN00154
 A.Experimental source: strain O157:H7, substrain RIMD 0509952
 C.Genetics:
 A.Gene: Ec93221

Query Match 3.9%; Score 109.5; DB 2; Length 879;
 Best Local Similarity 20.1%; Pred. No. 6.7;
 Matches 130; Conservative 68; Mismatches 208; Indels 241; Gaps 33;
 Oy 52 VSPSELTITDANLATPRLFLILGSPKRLTLAGSPAYLFGQTKDPLIPDKEST 111
 20 MSGSYVMAENBIQDPSRFLKGDFTKI-DLKRFSSGCVYEPG--KYNLVQVQNKQPLT 76
 Db 112 FEKSYQOVNODICKYGIIPDVEEKL-----RLEWYQQLREHYOKKFNSTYS 163
 77 EBYDIWYASNDASKTYVACTPRLVAVQFLKEDVARNLQWHDGKCLKKEQLE----- 130

QY 244 TYRSRSDVLYTPANCGLDLIFGH-NALLRTADLQNNSSN-----AQLLDYC----- 211
 DB 131 GIDIK-ADLSGALVLSLPOAVLEBTDINMPSPKMDGIGSLADYSITMOT 182
 QY 212 -----SSKGINI-SWEI-GNEPNSFL-KKADIFINGSO----- 241
 DB 183 RHEKGGDDSNISGNQGVNMGMRRLADQDTYLSKSNDDVDVINGDDTQKMEWSR 242
 QY 242 -----LGRDYIQLHKLKSTF-----KKAKYGGDV 268
 DB 243 YYAMRALPSLAKKDLGEDY-----LNSDIFGDNVYGGISITDDQMLPPLKGYADDI 296
 QY 269 GQPRKRTAKMLKSPKAGGEVI-----DSVTMHHYLYNGRTATREDFLN 312
 DB 297 SGVATTTAKVYVSQ--GRVYERQVAPGPRRIODLDGDSV-----SGTLHRIERON 346
 QY 313 PDVLDFISSVQKVPQVVESTRPKG-KWML-----GETS----- 345
 DB 347 GQVQEVYDINTASMP-----LTFPGQVRYKLMGPRQEWGHHVGGFSGGEGASWCIANGW 402
 QY 346 SAYGGAGLSD-----TFNA-----GFMWLDL-----GLSAR 374
 DB 403 SLVYGA--LDEHYQSAALGVGRDLSVFGAFAFDITSHTRLDKETAYGKGLDGNFR 459
 QY 375 MGI-----EVMWRQVFPAGVYHLVDENFDPLDYMILFEKLVGT---KVLASVQGS 426
 DB 460 LSYGDPPELMSRYTFAG---YRSEENFMMSER-LDASSEWRTDNDKEMTATATNQ 515
 QY 427 KRRKRLVYLCTNDNPRYKGGDLTLVAI-----NLNNTK----- 462
 DB 516 NFRDQGVVYLVNTHRTYWDREQTNVNMVLSHYNGLSIRMSISMGRYREYDNOAK 575
 QY 463 -YLRPFPSNKQVDKYLRLPLRPHGLSKSVQANGILTKMVDQT 507
 DB 576 GVTLSLSPMGDSSSTISY-----NGMYGSGSDSSQVG-YFSKVDAT 616

RESULT 10

S74760
 hypothetical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)
 C/Species: *Synechocystis* sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C/Accession: S74760
 R/Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 sp.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S74760
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-411 <KAN>
 A/Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PTDN:BA16911.1; PTD:dl01764
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 3.8%; Score 108.5; DB 2; Length 411;
 Best Local Similarity 21.4%; Pred. No. 2.5;
 Matches 66; Conservative 55; Mismatches 94; Indels 93; Gaps 17;
 QY 159 KFKASTYRSRSDVLYTPANCGLDLIF--GLNALRTADLQNNSSNAQL----- 206
 DB 151 KFRISPPREIDIT--FAGSTKLDLLASBENIDCIYHALANPRVYTSVAVAGQTLTMLRN 207
 QY 207 LLDYSSSGK--YNIWEL-----GNEPNSFLK-----ADIFINGSQGE 244
 DB 208 VIDVLADIDPIIYSSSWEIISGAGTTHADESTPALRGPERGKTYLAELI----- 260
 QY 245 DYIQLHLKLRSTFNANLKYGPVGQPRKRTAKMLKSGEVIDSVTWHHYLYNGRT 304
 DB 261 DHCRTRGLRCALRSSPVYSGMSDKP-----KFIENFKKASQOKQIVT--HHYING-- 311

QY 305 ATREDPLNPDV-----LDIFISSVQKVPQVVESTRPKGKWLGETSSAYGAGPLSDTFA 360
 DB 312 -----NPKDLHLHIDLSSIVATL-----KSRIGNINI-----GTGLSSTLK 351
 QY 361 AGFWMLDKLGLSA-----RKGIEVYMQVFPAGVYHLVDENFDPLDYMILFEKLVG 415
 DB 352 LAEMIRDELSSSMITQOIEVYVEVASIANNVGRAN-HVLD-----WEPVIFPS-QG 400
 QY 416 TKVLMASV 423
 DB 401 LKSLHQI 408

RESULT 11

D87541
 beta-xylosidase [imported] - *Caulobacter crescentus*
 C/Species: *Caulobacter crescentus*
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: D87541
 R.Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Bisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.O.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor
 n, J.; Rmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A/Reference number: A87249; MUID:21173698; PMID:11259647
 A/Accession: D87541
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-500 <STO>
 A/Cross-references: GB:AE005673; NID:g13423886; PTDN:AAK24328.1; GSPDB:GN00148
 C/Genetics:
 A/Gene: CC2357

Query Match 3.8%; Score 107.5; DB 2; Length 500;
 Best Local Similarity 25.7%; Pred. No. 4;
 Matches 56; Conservative 35; Mismatches 80; Indels 47; Gaps 15;

QY 165 YSRSDVLYTPANCGLDLIFGN--ALLRTAD--LQW--NSSNAQ-----LID-- 209
 DB 81 YDWTKIDQLYDALLAKGKIFELGFTPEAKMTSDQITFYWKGTSHPKLGPMRDLDIAF 140
 QY 210 -YCSSKGVNI-----SW--ELGNEPN--SPKKADIFINGSQLGEDYQLHKLKSTFKN 260
 DB 141 VHLIRARVGEVETWTFEWNENPNDGFWEKAD-----QAVYELDV-----TARA 188
 QY 261 AKLYGPD--VGQPRKRTAKMLKSF--KAGEVIDSVTWHHYLYNG--RTATREDPL 311
 DB 189 IKAIDFELRVGSPATAGAAVPEFLAHVKKSGSAVDVTHTYGVGDGFLDEKGVQDTKL 248
 QY 312 NPVDLDFISSVQKVPQVVE-STRPKGKWLGETSSAY 348
 DB 249 SPSP-DAVGDVDRVRROIBASAPGLPLVTEWSTSY 285

RESULT 12

F85875
 Probable fimbrial usher Z3600 [imported] - *Escherichia coli* (strain O157:H7, substrain E
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: F85875
 R.Penna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 11ler, L.; Grobbeck, B.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potemovskis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: F85875
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-879 <STO>
 A/Cross-references: GB:AE005174; NID:g12516702; PTDN:AGS7466.1; GSPDB:GN00145; UMG:Z360
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: Z3600

Query Match 3.7%; Score 106; DB 2; Length 879;
 Best Local Similarity 20.5%; Pred. No. 12;
 Matches 129; Conservative 64; Mismatches 201; Indels 234; Gaps 34;

52 VSPSFLSTIDANLADPPFLILGSPKLTARGLSPAYLFGGKTDFLIDPKKST 111
 20 MGSYVNAWAENBIQDSRFLEKGDPTKI-DLKRFSGGYVPG-KVNLQVQLNKOPLT 76
 112 PEERSWQGVNODICKSIPDVBEK-----RLEWYQBLLEHYQKKFKNS 163
 77 EBYDIWYASENDASKTYACLTPELVAQGLKEDVAKNLQWIDKCLKPGGLE----- 130
 164 TVSRSSVDLYTFANSGDLIFGL-NALLRTADLQWNSN-----AQLILDYC----- 211
 131 ----GIDIK--ADLSQALVSLPQALYLETIDIMDPPSMWDGSLGLIDYISITQOT 182
 212 ----SSKQYNI-SWEL-GNEPNSFL--KKADIFINGSO----- 241
 183 RHEENGDDNSNEISNGTGVNXXGAWRLPADWQTYLHSGSNDDVDVINGDDTQKNWMSR 242
 242 ----IGEDYIQLHKLKSTF-----KNAKLYGPDV 268
 243 YWAMRALPSLKAKGLGGEY-----LNSDIFGFNYGGSISTDDQMLPPVLRQYAPDI 296
 269 GQPRKRTAKMLKSLFKAGEVI-----DSVTMHHYILNGRTATEDFLN 312
 297 SGVATTAKVYISQL--GRVYETQVPAQPFRIQDLGDSV-----SGTLHIREON 346
 313 PDVLDFISSVQKVPQVVESTRPGK--KVWL-----GETS----- 345
 347 GQVQVEDINTASMPF-----LTRPGQVRKXLMGRPOEMGHVEGFFSGSGEASMGCIANGM 402
 346 SAYGGAPLLSD-----TFAA-----GFMLDL-----GLSAR 374
 403 SLVYGA--LADEHYQSAALGVGRDLVFGAVAFDITTSHTRLDKETAYGKGLDNGSFR 459
 375 MGI-----EVMRQVFFGAGNYHLVDENFDPLPDYWSLLPKLVGT--KVLMASVQGS 426
 460 LSYSDPFDLNRVTFAG--YRFSEENFTMSEY-LDMSDBEMRTGDKEMTYATITNQ 515
 427 KRRKLRYVLAHCTNTDNPYKESDGLTYALNLHNTKYTLR-----PYPSNK 473
 516 NFRDAGVSIVYLVNTRHTYHNDREQRYNNVL--SHYFVLGSIKMSISMGTGYREYDN- 571
 474 QVDK-----YLARPLGPHGLSKSVQLNG 497
 572 QADKGVYISLXMPMGD-----SSITSYNG 595

RESULT 13

T38446
 microtubule-associated protein smm4 - fission yeast (Schizosaccharomyces pombe)
 C.Species: Schizosaccharomyces pombe
 C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
 C.Accession: T38446; T00012
 R.McDougal, R.; Wood, V.; Barrrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, December 1997
 A.Reference number: Z21793
 A.Accession: T38446
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-670 <MCD>
 A.Cross-references: EMBL:AL009227; PIDN:CAA15832.1; GSPDB:GN00066; SPDB:SPAC27D7.13C
 A.Experimental source: strain 972h; cosmid c27D7
 R.Yamashita, A.; Matanabe, Y.; Yamamoto, M.
 Genes to Cells 2, 155-166, 1997
 A.Title: Microtubule-associated coiled-coil protein smm4 is involved in the meiotic deve
 A.Reference number: Z14042; MUID:9731155; PMID:9167972
 A.Accession: T00012
 A.Status: translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-670 <YAM>

A.Cross-references: EMBL:AB000269; NID:93341860; PIDN:BA31857.1; PID:93341861
 C.Genetics:
 A.Gene: smm4; SPAC27D7.13C
 A.Map position: 1

Query Match 3.7%; Score 105; DB 2; Length 670;
 Best Local Similarity 21.4%; Pred. No. 9.6;
 Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15;

107 KKSTFEERSWQGVNODICKSIPDVBEKRLLEWYQBLLEHYQKKFKNSTYS 166
 73 KRPSVNSR-----KKGSEINSMFEKTKALKQKSRREPSKFEKSLAPLCITPIDSTPT 128
 167 RSSVDVLYTFANSGDL-LIFGLNALLRTADLQWNSNAOLL-LDYCSSKQYNI-SWELGN 224
 129 KTA--TFTYSTTENDELINFSTBELSSPOTTLNLSSTLSGLDSSFMEEBEVWQVDN 186
 225 ----BPNSEFLK-----ADIFINGSQLGEDYIQLHKLK--ST 257
 187 VLQCEKFKTPHSGSYLKENLSLRKRLDELIMCENTALKEKIDLNKELEKVEPOLT 246
 258 FKNAKLYGPDVQGR--KTAKMLKSLFKAGEV-----DSVTMHHYILNGRT 304
 247 FLRSK--NSIEKRNFRREKFLKFLAMQKEIKYLRRLQIRKIDIPYKISDRSLSKT 303
 305 ATREDPLNPVDL--LFISSVQKVPQVVESTRPGKVMGETSSAYGAGAPLLSDTFAA 361
 304 PKSDNMNTQVTPSSLLGVEVSRYDL-----KQVQDYTF----- 340
 362 GFWMLDKLGASARKGIEVMRQVFFGAGN--YHLVDENFDPLPDYWSLLPKLVGTK 417
 341 -LVKIPKPNPSEKLTISNVRYLIVPGSLDLQSLTNEFN--VHMNSTVYQELNLTK 395
 418 VLASVQGSRRK 430
 396 SNNSVDGVKTR 408

RESULT 14

S00652
 phosphoribosylamine-glycine lyase (EC 6.3.4.13) - fission yeast (Schizosaccharomyces po
 N.Alternate names: AIRase; aminomidazole ribotide synthetase; GARase; glycylamide rib
 N.Contains: phosphoribosylamine-glycine lyase (EC 6.3.4.13); phosphoribosylformylglycin
 C.Species: Schizosaccharomyces pombe
 C.Date: 07-Sep-1990 #sequence_revision 28-Oct-1994 #text_change 03-Jun-2002
 C.Accession: S00652; T40496; T40422
 R.McKenzie, R.; Schuchert, P.; Kibbey, B.
 Curr. Genet. 12, 591-597, 1987
 A>Title: Sequence of the bifunctional ade1 gene in the purine biosynthetic pathway of th
 A.Reference number: S00652; MUID:89003164; PMID:3502942
 A.Accession: S00652
 A.Molecule type: DNA
 A.Residues: 1-788 <MCK>
 A.Cross-references: EMBL:X06601; NID:94903; PIDN:CAA29820.1; PID:94904
 R.Wood, V.; Rajandream, M.A.; Barrrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.
 submitted to the EMBL Data Library, February 1998
 A.Reference number: Z21910
 A.Accession: T40496
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-788 <WOO>
 A.Cross-references: EMBL:AL021730; PIDN:CAA16823.1; GSPDB:GN00067; SPDB:SPBC4C3.02C
 R.Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrrell, B.G.
 submitted to the EMBL Data Library, March 1999
 A.Reference number: Z21928
 A.Accession: T40422
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 604-788 <SE>
 A.Cross-references: EMBL:AL035555; PIDN:CA38600.1; GSPDB:GN00067; SPDB:SPBC405.01
 A.Experimental source: strain 972h; cosmid c405
 C.Genetics:

A:Gene: A081; SPDB:SPBC405.01

C:Map position: 2

C:Superfamily: Saccharomyces cerevisiae ADE5 multifunctional protein; phosphoribosylamin

C:Keywords: cyclo-ligase; purine nucleotide biosynthesis

F/5-425/Domain: phosphoribosylamine-glycine ligase homology <PGH>

F/439-767/Domain: phosphoribosylformylglycinamide cyclo-ligase homology <PFCL>

Query Match 3.7%; Score 104.5; DB 1; Length 768;

Best Local Similarity 27.7%; Pred. No. 13;

Matches 70; Conservative 36; Mismatches 114; Indels 33; Gaps 11;

297 HYVNGRTATRE--DPLNDV-LDIFISSVQVQVVEST-RPGKVMLETSSAAY---- 348

424 HHALNPKRKRREITLTVSSGVSVNGNEFVORIDLVKSTRPADADIGFGFGLPLQ 463

349 -GGGAPLL-SDTFAAGFMWLDKLGLSAR--MGIEVNRQVFFGAGNYHLYDENPDL-P 402

484 AGNNDPLVSVATDGVSKLIALSLNKHDTVGIDLVAMNV-----NDLVVQGAELIFL 537

403 DYMLSLPKLVGKVLMAVQSGSKRRKRVYLCTNDNPRYKSGDLTLVAINLHNVTK 462

538 DYFATGSLDLKVSSTFVGVVKGCKGACALVGETSEMPGLYHDGHYDANGTSVGAVSR 597

463 YLRVPYPSNKKQVQVLYLRPLGPHGLSKSYQVANGTLT--KQVD-----DOTLPPLMEKPL 516

598 DDLPKRESFSKIDIL-----GLASDGVHNGSYLVKRIKIVSDLETTSVCPMDKNV 650

517 RPSGLPLPAPSY 529

651 RLGDLSLIPTRY 663

RESULT 15

adenylosuccinate synthetase - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999

C:Accession: F70411

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: F70411

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-432 <AOF>

A:Cross-references: GB:AE00733; NID:92983720; PIDN:AAC07286.1; PID:92983724; GB:AE00065

A:Experimental source: strain VFS

C:Genetic:

A:Gene: pura

C:Superfamily: adenylosuccinate synthase

Query Match 3.7%; Score 104; DB 2; Length 432;

Best Local Similarity 23.9%; Pred. No. 5.9;

Matches 96; Conservative 39; Mismatches 128; Indels 138; Gaps 22;

15 LLLGLPLSPGALPPAQADVVDLD-----FTQEPHLVSPS 55

51 ILHLPLPGILHBNHKGVAQGM-VVDLEVLHKEVKNLEBKGIYKERLFISDRALHMPY 109

56 PLAVTIDANLATDPRFLILGSPK-LRTIARGLSPAYL-RFGTKTDFLIDPKKESTP 112

110 H-----KLDSLFEKKKGIGITLKGIGAPMPKYG--RKGIKISDLKDEKRF 154

113 EERSYMQSVQVODICKYGIIPDVEK-----LRLEMPQEQOLLREHQQKKKXSTY 165

155 -----YTLIEDNLDVFK-----NICEKVFCEKFDIDINQIYEQL-----RYEEFKENV- 199

166 SRSSVDVLYTPANGSGDLIFGLNALRTADL-----QNNSSNAQLLDYCSSKGYNISWE 221

200 -----VDLIRFNTQKGSVLFEGAGQTLIDVDMGTVPYVTSNASAL-----GLSNG 246

QY 222 LGNEPNSPLKADIFING-----SOL-GEVYIQLHKLKSTFKNAKLYG 265

Db 247 TGNPPTKTF--SDAFPLGVAKAYTRVGGSPPTTEKGEKREL-----GGEYG 295

QY 266 PDVGQPRR--KTAKMLKSPLAGGEVIDSVTWHHYLYNGRTATREDPLNP----- 313

Db 296 STGRPRRCGMIDLVLKVAQVNG-----LDGFVITKLDVDTFDEVKVCVA 343

QY 314 -----DVLDFISSVQVQVQV--VESTPRGKVMLETSSA 347

Db 344 YELDGEVIDYFPASYSSELIRKRVVYKTLKG---WKKSTGCA 381

Search completed: October 22, 2003, 20:27:38

Job time: 46 secs

QY 227 NS-----FL-----KKADIFINGSQLGDEHYTOLHNLKSTF 258
 DB 110 NAEIDWVEICEBVLFLFLFDIDPYKEHEIIKOKVREFFLN-KIGIEY-----KSDY 159
 QY 259 KNAKLYGPDVGQPRRTAKLKSFLKAGGEVIDSWTHHYTLNGRTAT-REDFLNPDV 316
 DB 160 DNINLV---VGDAKRFIKSDKXY-----NVVPHDAPSPKRPDTLYTDFL-----202
 QY 317 DISSVQKVFQVVESTRPKKVMGLGTSAYGGAFLSDTFAAGFMMLDKGLSABMG 376
 DB 203 -----KEIKYKMEBN--CVLI-----SYSSALPFSALVDGCFVISEKSVGRKGG 246
 QY 377 IEVWMEQVFAGNHYHLDVENDP-----PLPDYMLSLFLKLVGTYKVLMAVSGSKR 429
 DB 247 ITLAYKNPNKPRINVEDERYALSYIALPYRDETLISLKKIIDREPRREKLEKLI 306
 QY 430 KAVYTLHCTNTDNPRIKYGDLTYA--INLHNTKYLRLPY 468
 DB 307 KIGKYLSTKQIKKKNPEELIKLOKEDLNSSEIKKWRLEK 347
 RESULT 2
 YB9F_YEAST
 ID YB9F_YEAST STANDARD; PRT; 688 AA.
 AC P38338;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical 80.4 kDa protein in POP4-SH1 intergenic region.
 GN YBR235W OR YBR1727.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NC NCBI_TaxId=4932;
 RX MEDLINE=93220397; PubMed=8465606;
 RA Dolignon P., Bileau N., Crouzet M., Aigle M.;
 RT "The complete sequence of a 19,482 bp segment located on the right
 arm of chromosome II from Saccharomyces cerevisiae."
 RL Yeast 9:189-199(1993).
 CC CC
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 DR EMBL; X70529; CAA49923.1; -;
 DR EMBL; Z36128; CAA85222.1; -;
 DR PIR; S32961; G32961.
 DR SGD; S0000463; YBR259W.
 KW Hypothetical protein.
 KM SEQUENCE 688 AA; 80426 MW; 0BA84837BD7A4B30 CRC64;
 SO
 Query Match 3.9%; Score 111; DB 1; Length 688;
 Best Local Similarity 22.5%; Pred. No. 1.6;
 Matches 67; Conservative 45; Mismatches 94; Indels 92; Gaps 16;
 QY 126 ICYKGSIPPEVERKRLLEMYOQOLLREHYOKKFKSTYSRS-----VDVLT 175
 DB 164 MAATSSKRWSDDRQLOFMFEPFMKKECLVKEFVFDLQKSSDPLKELIIMWEKIVV 223
 QY 176 FANCSGLDLIFGNALRTADLOMNSN-----AQLILD-----YCCKGY- 216
 DB 224 -ANC--IDATGQVRIDGAEIIMTSKNLVFSSISAVLRLNLONNFSAFRYGEALV 280
 QY 217 -----NISWEIGNEBNSFLKKA--DIF--INGSLG--EDYIOLHKLRLK-----255
 DB 281 QDEPAHRSLEKMDSDKVESILRALIFNDMEFPYFKKEQVDTKADGIFELRLLRKFKEHIN 340

QY 256 -----STFNK--AKLYGPDVGQPRRTAKMLKSFLKAGGEV-----IDSV 293
 DB 341 DYKDHQIVIKIANSQFANNSTLMTSSKTDPRKSNMPSILDDGKIKGMHVSPIDE- 399
 QY 294 TWHHYTLNG-----RTATREDFLNPDVLDIFISSVQKVFQVVESTR---PGK 338
 DB 400 -YSHFIDNDEPLMDKVPKIKYTNQOTPTPDASAFIDS--HKIYALISLRYVLPBK 454
 RESULT 3
 TEXT EUPAE
 ID TEXT EUPAE STANDARD; PRT; 1031 AA.
 AC 000939;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Telomerase reverse transcriptase (EC 2.7.7.-) (Telomerase catalytic
 DE subunit) (telomerase subunit p123).
 OS Euplotis aedicularis.
 OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
 OC Euplotida; Euplotidae; Euplotes.
 NC NCBI_TaxId=5940;
 RX MEDLINE=97274210; PubMed=9110970;
 RA Lininger J., Hughes T.R., Shevchenko A., Mann M., Lundblad V.,
 RA Cech T.R.;
 RT "Reverse transcriptase motifs in the catalytic subunit of
 RT telomerase."
 RL Science 276:561-567(1997).
 CC CC
 CC CC -1- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR
 CC THE REPLICATION OF CHROMOSOME TERMINI IN MOST EUKARYOTES. IT
 CC ELONGATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT ADDS
 CC SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE
 CC SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.
 CC CC
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 DR EMBL; U95964; AAC47515.1; -;
 DR InterPro; IPR000477; RYase.
 DR InterPro; IPR003545; Telomerase_RT.
 DR Pfam; PR00078; rvt; 1.
 DR PRINTS; PR01365; TELOMERASERT.
 KW Telomerase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
 KW DNA-binding.
 KM SEQUENCE 1031 AA; 122562 MW; 57B87A63A1FED60F CRC64;
 SO
 Query Match 3.7%; Score 105.5; DB 1; Length 1031;
 Best Local Similarity 20.9%; Pred. No. 7.5;
 Matches 81; Conservative 56; Mismatches 133; Indels 117; Gaps 17;
 QY 102 LIFDPKESSTFEERYSQVQVODICKYGSIPPDVEBKRLLEMYOQOLLREHYOKKFK 161
 DB 694 LIVEAKQRYVFKKNDLQPVIN--ICQNYI-----NFGKGFY 729
 QY 162 NST-----YSSRSVDVLYTFANCSGLDLIF-----GLNALRTAD-----LOMN 200
 DB 730 KQTKGIPQGLCVSSILSSFFYYATLEESSIGFLRDESNMPPNNVLLKRLDDYLLITQ 789
 QY 201 SSNAOLLDDYSSSKRYNISWEIGNE-----PNSFLKADAFINGSQLGDEHYI 247
 DB 790 ENNAVLFLE---KLINVSRENGFEPNMKKIQTSPSPSPFAKYGMDSVBEQNIVDYIC 845

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC -----
CC EMBL, AB000269, BAA31857.1,
CC DR EMBL, AL009227, CAA15832.1,
CC DR EMBL, AL034583, CAA22580.1,
CC PIR, T38446, T38446,
CC DR GeneDB, Sprobe, SPAC2707.13c,
CC InterPro, IPR000938, CAP-Gly,
CC DR Pfam, PF01302, CAP_GLY_1,
CC DR PROSITE, PS00845, CAP_GLY_1,
CC DR PROSITE, PS50245, CAP_GLY_2,
CC DR Meiosis, Microtubules, Coiled coil,
CC KW DOMAIN 23 65 CAP-GLY
CC FT DOMAIN 209 254 COILED COIL (POTENTIAL)
CC FT DOMAIN 404 582 COILED COIL (POTENTIAL)
CC SO SEQUENCE 670 AA, 77105 MW, 011BD740211B57C CRC64;
CC
CC -----
CC Query Match 3.7% Score 105; DB 1; Length 670;
CC Best Local Similarity 21.4%; Pred. No. 4.4;
CC Matches 80; Conservative 58; Mismatches 149; Indels 86; Gaps 15
CC
CC QY 107 KKESTFEERYSWQSVQVNDICKYGIIPDVEEKLRLMPYQEQLLREHYQKKNSTYS 166
CC Db 73 KRPSVVSFR---KKSESNISNFEKTKAIKQSRREPSKERSLARDLCITPIDSSPT 128
CC QY 167 RSSVDVLYTFANGSGLD-LIFGLNALRTLDLQWSSNAQLL-IDYSSKGYNISWEIGN 224
CC Db 129 KTA--FTYTSSTTEMDLNFSTTELSFDTLLNSDTSKLSGDDSSFMEEEFVQVND 186
CC QY 225 -----EENSFLTK-----ADIFINGSQLEGDYIOHLKLRK---ST 257
CC Db 187 VLQCEKKFLPHSGKSTLKENLSELRKGRDELMLCENTALKEKIDKLANKLEKVEPOL 246
CC QY 258 FKNAKLYGPPVGQPRR-KTAKMLKSLFKAGGEV-----IDSVTHMYHYLNGRT 304
CC Db 247 FLNRK--NSLEKRNRRREKELKKFLAMQKEIKYLRKRKLQIRKIPYKYSDRSLNKT 303
CC QY 305 ATREDFLNPVLD--LFISSVQKVFQVVESTRPQKMYLGETSSAYCGGAPLSDTPAA 361
CC Db 304 PKSGDNMTQVTPESLLGVSFVSRYLDL-----KQYQVADITE----- 340
CC QY 362 GFMMWLDLGLSARNGIEVNRQVFFGAGN----YHVDENDPDLPDVWLSLFLKLVGTK 417
CC Db 341 -LVVPIKPNPSEKLTJNNVNRKYLNIIVGSLDLPSTLNENF---VHNNSTVYQELLNWK 395
CC QY 418 VLMAVQSGSKRK 430
CC Db 396 SNNSVDGAKTR 408
CC
CC -----
CC RESULT 5
CC THIC.VIBPA STANDARD; PRT; 646 AA.
CC AC 087KF0,
CC DT 15-SEP-2003 (Rel. 42, Created)
CC DT 15-SEP-2003 (Rel. 42, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Thiamine biosynthesis protein thic.
CC GN THIC OR VP3027.
CC OS Vibrio parahaemolyticus.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
CC OC Vibrionaceae; Vibrio.
CC OX NCBI_TaxID=670;
CC RP [1]
CC SEQUENCE FROM N.A.
CC RC STRAIN=RTMD 2210633 / Serotype O3:K6;

```

RX MEDLINE=22508454; PubMed=12620739; Yokoyama K., Uda T., Tagomori K.,
 RA Makino K., Oshima K., Kurokawa K., Kubota Y., Yamashita A., Kubota Y., Kimura S.,
 RA Iijima Y., Naito M., Naito M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shingawa H., Hattori M., Iida T.,
 RT Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 RT distinct from that of *V. cholerae*.
 RL Lancet 361:743-749(2003).
 CC -1- FUNCTION: Required for the synthesis of the hydromethylpyrimidine
 CC (HMP) moiety of thiamine (4-amino-2-methyl-5-
 CC hydromethylpyrimidine) (By similarity).
 CC -1- PATHWAY: Thiamine biosynthesis.
 CC -1- SIMILARITY: BELONGS TO THE THIC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC EMBL: AF005083; BAC61290.1; -
 DR HAMAP; MF 00089; -1.
 KW Thiamine Biosynthesis; Complete proteome.
 SQ SEQUENCE 646 AA; 72477 MW; FED3B87A88C237 CRC64;
 Query Match 3.7%; Score 104.5; DB 1; Length 646;
 Best Local Similarity 19.4%; Pred. No. 4.5; Indels 147; Gaps 23;
 Matches 86; Conservative 56; Mismatches 153;
 QY 43 FTQEPRLHVPSPFLSVITDANLAPRPLILH-GSPKLR-----TLARLSA 90
 DB 61 FEENEPRVYDTS-----GYTDPDVAIDLYSGPLKREGMIEERDTEILDVSSV 112
 QY 91 Y-----LRGG-----TKTDL--IFPKKESFEESYQSV 122
 DB 113 YAKRLDDETLDDLRYGNLPRIRARAKGCVTLAHYARKGIVPEMEYIALREMGRAQY 172
 QY 123 NODIC-----KYG-SIPDVEKL-----RLMEPQOELLREHYOKFE 160
 DB 173 RQDVLTQGHGSGFANLPDITAEPRKREKAVAGRAIIPSNINHPSEPMITIRNLVAV 232
 QY 161 K-----NSTYSRS--SYDVLTPFANCSG--LDLIFGLNA-----LRLTADQWNSNA 204
 DB 233 NANIGNSSVTSIIEBEVEKLWATRCGDTWDLTSGRNHETREWLNSPVPICVTVM 292
 QY 205 QLLLDYSSSGYVISMELNEPNSFLKKADIFNSQGLGEDIYQLK--LKRSTKNAK 262
 DB 293 YQALEKNGIAEMINWEV-----MRDTLLEQAEQGVDTYTHAGLLR----- 335
 QY 263 LVGPDVGPARRKAKMLKSFLLKAGGEVDSVW--HHYYLNGRTATREDPLNPDLVI- 318
 DB 336 -YVP-----MTAKRYTVISRGSGSI--AKKCLAHN-----QSFLLYTHREIC 376
 QY 319 -FISSVQKQVQVESRPRGKRWLGETSSAYGGAPLSDTPAAGFMWLDKGLSARWGI 377
 DB 377 EICAKTDVALSIDGRLRPGS-----VADANBAQAFALRTIGELFKIAM 420
 QY 378 EYVMROVFPAGN--YHVDENED 399
 DB 421 EYDVOVIEEPGHIPIHMIKEND 444
 RESULT 6
 PUR2_SCHPO STANDARD; PFT; 788 AA.
 ID PUR2_SCHPO
 AC P20772; Q9UM5;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bifunctional purine biosynthetic protein Ade1 [includes: (Glycinamide
 DE phosphoribosylamine-glycine ligase (EC 6.3.4.13) (GARS) (Glycinamide
 DE ribonucleotide synthetase) (phosphoribosylglycinamide synthetase);

DE Phosphoribosylformylglycinamide cyclo-ligase (EC 6.3.3.1) (AIRS)
 DE (phosphoribosyl-5-aminimidazole synthetase) (AIR synthetase)].
 GN Ade1 OR SPBC405.01 OR SPBC43.02C
 OS Schizosaccharomyces pombe (fission yeast).
 OC Burkholderia; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes
 OC NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=89003164; PubMed=3502942;
 RA McKenzie R., Schuchert P., Kilbey B.,
 RT "Sequence of the bifunctional ade1 gene in the purine biosynthetic
 RT pathway of the fission yeast *Schizosaccharomyces pombe*,"
 RL Curr. Gene. 12:591-597(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymopoulos B.,
 RA Welford I., Vansireels E., Rieger M., Schefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinbach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode J.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.,
 RT "The genome sequence of *Schizosaccharomyces pombe*,"
 RT Nature 415:871-880(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + 5-phospho-D-riboylamine + glycine = ADP
 CC + phosphate + N(1)-(5-phospho-D-riboyl)glycinamide.
 CC -1- CATALYTIC ACTIVITY: ATP + 2-(formamido)-N(1)-(5-phospho-D-
 CC ribosyl)acetamide = ADP + phosphate + 5-amino-1-(5-phospho-D-
 CC ribosyl)imidazole.
 CC -1- PATHWAY: De novo purine biosynthesis; second step.
 CC -1- PATHWAY: De novo purine biosynthesis; fifth step.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AIRS FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AIRS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC EMBL: X06601; CA229820.1; -
 DR EMBL; AL021730; CA16823.1; -
 DR EMBL; AL035655; CAB38600.1; -
 DR PIR; S00652; S00652.
 DR HSSP; P08178; ICL1.
 DR GeneDB; Spombe; SPBC405.01; -
 DR InterPro; IPR000728; AIRS-related.
 DR InterPro; IPR000115; Gars.
 DR InterPro; IPR004733; Purm_c1lgase.

DR Pfam; PF00586; AIRS; 1.
 DR Pfam; PF02769; AIRS C; 1.
 DR Pfam; PF01071; GARS; 1.
 DR Pfam; PF02842; GARS; 1.
 DR Pfam; PF02843; GARS C; 1.
 DR Pfam; PF02844; GARS N; 1.
 DR TIGRPFAM; TIGR00877; purM; 1.
 DR TIGRPFAM; TIGR00878; purM; 1.
 DR PROSITE; PS00184; GARS; 1.
 DR Multifunctional enzyme; Purine biosynthesis; Ligase.
 FT DOMAIN 1 430 GARS
 FT DOMAIN 1 430 GARS
 FT DOMAIN 1 430 GARS
 SQ SEQUENCE 788 AA; 85231 MW; 0FDE64REAF9095D CRC64;
 Query Match 37%; Score 104.5; DB 1; Length 788;
 Best Local Similarity 27.7%; Pred. No. 6;
 Matches 70; Conservative 36; Mismatches 114; Indels 33; Gaps 11;
 297 HYLLNGRTATRE--DELNDPV-LDIFISGVKQVQVVEST-RPGKKVNLGETSSAV--- 348
 424 HPAALPKRTREILTYENGSGVNDNGNEFVQIKDLVSTRPGADADIGGCGIFDLKQ 483
 349 -GGAPPL-LSDTFAGFMWLDKLGSLAR--WGIEVVMRCVFPFGAGNYHLVDENFDPL--P 402
 484 AGMNDPLVASATDVGSKLLIALSLNKHDTGIDLVANV-----NDLVGGAELIFL 537
 403 DYWLSLFPKLVGTGYKVLASVQSGSKRRKRVLYLCTNTNPPRYKSGDLTVAINLHNTK 462
 538 DYFANGSLDLKYSTFVGVGVKQKQAGALVGETSEHPRGLYHGHYDANGSTVCANVR 597
 463 YLRLPFPAENKQVDKYLRLPGLPHGLSKSYOLNGTLT--KNVD---DQTLPLMEKPL 516
 598 DDLRKPSFSGKIDLL-----GLASDGVHNSNGSLVTKIVTSIDLYTSVCPMDKNV 650
 517 RPSGSLGLPAPSY 529
 651 RLGDLSLIPRIY 663
 RESULT 7
 PURA_AOUAE STANDARD; PRT; 432 AA.
 ID AC 067321;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DT Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase) (AdSS) (AMPSase).
 DE PURA OR AQ.1290
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 NC NCB1_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5.
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.B., Overbeek R., Ostead M.A., Keller M., Anjay M., Huber R.,
 RA Feldman R.A., Short J.M., Sison G.J., Swanson R.V.;
 RA "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus".
 RL Nature 392:353-358 (1998).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE
 CC NUCLEOTIDE BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-aspartate = GDP + phosphate +
 CC adenylosuccinate.
 CC -1- PATHWAY: AMP biosynthesis; first committed step.
 CC -1- SIMILARITY: Belongs to the adenylosuccinate synthetase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ldb-sib.ch/announce/>
 CC or send an email to license@ldb-sib.ch).
 CC
 DR EMBL; AE000733; AAC07286.1; -
 DR PIR; F70411; F70411.
 DR HSSD; P12283; IADE.
 DR HAMAP; MF_00011; -.
 DR InterPro; IPR001114; Aducc synthetase.
 DR Pfam; PF00709; Adenylosucc_synth; 1.
 DR ProDom; PD001188; Aducc synthetase; 1.
 DR TIGRPFAM; TIGR00184; purA; 1.
 DR PROSITE; PS01266; ADENYLOSUCIN SYN 1; 1.
 DR PROSITE; PS00513; ADENYLOSUCIN SYN 2; 1.
 DR Purine biosynthesis; Ligase; GTP-binding; Complete proteome.
 FT NP BIND 12 18
 FT ACT SITE 137 137 BY SIMILARITY.
 FT ACT SITE 144 144 BY SIMILARITY.
 SQ SEQUENCE 432 AA; 48854 MW; ACA52CE913DB62ET CRC64;
 Query Match 37%; Score 104; DB 1; Length 432;
 Best Local Similarity 23.9%; Pred. No. 2.8;
 Matches 96; Conservative 39; Mismatches 128; Indels 138; Gaps 22;
 15 LLLGEPPLSPGLPRPAQADVDLD-----FTQEPHLVPS 55
 51 ILHLPLTILHEHVKVYAOQM-VVDLEVLKENVKEKGIYKERLFISDRHVLNMPY 109
 56 FLVYTDANLMTDPRFLILGSPK--LRTLRGLSPAYL-RFGGKTDFLIFDKBESTF 112
 110 H-----KLDLSLEKKKGIGITLRGIPAPMFYRG--RKQIRISDLKDEKRF 154
 113 EBRSTWQVQNDICKGSIPTDYBEK-----LRLEMPQOQLREHQQKPKYSTY 165
 155 ---YTLLEDMLDPVK-----NICEKVFCEKFDLDINQIYEOL---RFESEKENV- 199
 166 SRSSVDVLYTFRANGSGDLIRGNALLRTADL-----QWSSNAQLLDYCSKGVNISM 221
 200 ---VDLRFNPTOKGSVLFEGAOCTLDVDMGYTPYVTSNNAAL-----GLSNG 246
 222 LGNEPNSFLKRDIFING-----SQL-GEYIOLHLKLRKSTPRNAKLYG 265
 247 TQMPKRYF---SDAPFLGVAKAYTTRVGECPFLTEKGESEBKRLREL-----CGEYG 295
 266 PDVGQPRF--KTAAMKLSFLKAGEVYDVTMHYIYNGTATREDPLN----- 313
 296 STTGPRRCGMDLVALKYAVQVNG-----LDGVIITKLDVLDTFDEKVCVA 343
 314 ---DVLDFISSVQKFOV--VESTRPKKWLGETSSA 347
 344 YELDGEVIDYFPASISELIRKVPYKTLKG--WKSTXGA 381
 RESULT 8
 RRP2_IAKIT STANDARD; PRT; 716 AA.
 ID AC 091742;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase subunit P2 (EC 2.7.7.48) (Polymerase
 DE acidic protein) (PA).
 OS Influenza A virus (strain A/Kitakyushu/159/93).
 OC Influenza; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza A viruses; Influenzavirus A.
 NC NCB1_TaxID=62478;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94406205; PubMed=9733841;
 RA Lindstrom S.B., Hiromoto Y., Nerome R., Omoe K., Sugita S.,
 RA Yamazaki Y., Takahashi T., Nerome K.;
 RA "Phylogenetic analysis of the entire genome of influenza A (H3N2)
 RT viruses from Japan: evidence for genetic reassortment of the six

CC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
CC	Bacterium
CC	NCBI TaxID=1489;
CC	SEQUENCE FROM N.A.
CC	STRAIN=ATCC 824 / DSM 792 / VGM B-1787;
CC	MEDLINE=21359325; PubMed=11466286;
CC	Moelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
CC	Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,
CC	Tatunov R.L., Sabath F., Doucette-Stamm L., Soucaille P., Daly M.J.,
CC	Bennett G.N., Koonin E.V., Smith D.R.
CC	"Genome sequence and comparative analysis of the solvent-producing
CC	bacterium Clostridium acetobutylicum";
CC	J. Bacteriol. 183:4823-4838(2001).
CC	-1- COFACTOR: Thiamine pyrophosphate (potential).
CC	-1- SIMILARITY: BELONGS TO THE XFP FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@isb.ch).
CC	EMBL; AE007645; AAC79311.1;
CC	PIR; D97065; D97065.
CC	HMAP; MF 01403; -1.
CC	Interpro; IPR000399; Pyruvate_decarb.
CC	Interpro; IPR005593; XFP.
CC	Pfam; PF03894; XFP; 1.
CC	PROSITE; PS60002; PHOSPHOKETOLASE 1; 1.
CC	PROSITE; PS60003; PHOSPHOKETOLASE 2; 1.
CC	PROSITE; PS00187; TPP ENZYMES; FALSE NEG.
CC	KM Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
CC	SEQUENCE 796 AA; 90640 MM; 488219DC778FAAE CRC64;
SO	
SO	Query Match 3.6%; Score 103; DB 1; Length 796;
SO	Best Local Similarity 21.6%; Pred. No 7.9;
SO	Matches 77; Conservative 51; Mismatches 114; Indels 114; Gaps 18;
QY	27 GALPRQAQDVLDLFTFOEPHLVSPSLVITDA---NLANDPR-LILGSPKRT 82
DB	238 GWKYEYFEGEDPEPETHKLMATLTDIVIEELINICKARENNDCSRPKPMIVLRTPK--- 294
QY	83 LARGLSPAVILRPGSTKTDPLIFDPKKESTFEERSYQVQNDICKYGIIDPVEKRL 142
DB	295 ---GVTGPKFV-----DGVNKGSPRAQVLAVDRIYTEMDDLE----- 332
QY	143 EW--PYOEQLLRREHYO--KKFKNSTYRSRSVDLYTFANCSGLDIFGLNALRTADLQ 198
DB	333 EMLKSYRPEELPDENVRLILPELBELTLPKGNRMANLHAN--GGL-----LRLRLTPDR 386
QY	199 WNSSNAQLLDYCCSSKGYNISWELGNEPNSFLKPADIFINGSQIGEDYQLHKLR----- 254
DB	387 -----DYA-----VDVPTPGSTVKQMDIEIGKTVRDVVK 415
QY	255 -KSTFKAKIYGP-----VGQPRKTKAKLK--SFLKAGGEVDSVTWH-- 297
DB	416 LNEEDTRNFRIRFGPELTSNRMLAVFEGTKQWLSLIEPNDDEFSLNDRIVDSMISELIC 475
QY	298 -----YLLNGRTATREDPLNPDVLDIFISSVQKFOVES--TRPGKRWIGETS 345
DB	476 EGMLEGLILVGRHG-----FPASVEAFNLIVDSMITQHGK--WLKVTYS 516

DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
 DE (MeRS)
 GN MENS OR CAC2991.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxId=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Brecon G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,
 RA Tatusov G.N., Sabathe F., Doucellet-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.,
 RA "Genome sequence and comparative analysis of the solvent-producing
 RA bacterium Clostridium acetobutylicum."
 RL J. Bacteriol. 183:4823-4838(2001)
 CC -1- FUNCTION: Is required not only for elongation of protein synthesis
 CC but also for the initiation of all mRNA translation through
 CC initiator tRNA (fMet) aminoacylation (By similarity)
 CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
 CC L-phosphatate + L-methionyl-tRNA(Met)
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC MeCG subfamily 2A.
 CC -1- SIMILARITY: Contains 1 tRNA-binding domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL, AB007796; AAK80932.1; --
 DR PIR, A97268; A97268.
 DR HAMAP, MF_01228; fused; 1.
 DR InterPro: IPR004495; MeCG_Cterm.
 DR InterPro: IPR002300; tRNA-synt_1a.
 DR InterPro: IPR001412; tRNA-synt_1.
 DR InterPro: IPR002304; tRNA-synt_mec.
 DR InterPro: IPR002547; tRNA bind.
 DR Pfam, PF00133; tRNA-synt_1; 1.
 DR Pfam, PF01588; tRNA bind; 1.
 DR Pfam, PF001528; MeRS dimerising; 1.
 DR PRINTS, PRO1041; TRNASYNTHET.
 DR TIGRFAMs, TIGR00398; mecg; 1.
 DR TIGRFAMs, TIGR00399; mecg_C_term; 1.
 DR PROSITE, PS00178; AA tRNA_LIGASE_I; 1.
 DR PROSITE, PS50886; TRSD; 1.
 KW RNA-acyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
 KW RNA-binding; tRNA-binding; Metal-binding; Zinc; Complete proteome.
 FT SITE 14 24
 FT SITE 29 303
 FT DOMAIN 542 644
 FT METL 129 129
 FT METL 132 132
 FT METL 146 146
 FT METL 149 149
 FT BINDING 302 302
 FT SEQUENCE 644 AA; 73618 MW; 30035F8B94B60A7A CRC64;
 Query Match 3 6%; Score 101; DB 1; Length 644;
 Best Local Similarity 18.4%; Pred. No. 8.2; indels 145; Gaps 19;
 Matches 78; Conservative 58; Mismatches 145; indels 142; Gaps 19;
 163 STYSSRSVDVLTYPANCGLDIFGIALMLRTADLQWNSNNAQLLDYSSKGVNIS--- 219
 24 NVTYTVASALVRFRLQGYDAFM-----LGTGDEHQKQRLAABDGITPKAVV 73

QY 220 -----WELGN-EPNSFLKKADIFINGSLGEDIYOLHKLKSTFKNALYG--- 265
 DB 74 DEIVAGIKDKMKNMNIYDFIRTTD-----EHVKAQVQIKVKKFDNGIYISAY 124
 QY 266 -----PDVGQPRRKRTAKMLKSF--LKAGEVIDSVTHMYLYN 301
 DB 125 EGWYCTPCESEFWETQVLVDGKCPDGRPVETKEAEVFKMSKADRLIKYIEHPPIQ 184
 QY 302 GRITRE--DFLNPVDLDFISSVQVQVYESTRGKKWLGETSAYGGAPILSDT 358
 DB 185 PESKRNEMLNFLRPGIQLDCLIS-----SSSPMGPIPFDE 221
 QY 359 FAAGFMWLDKI-GLSAMGLEVWVROVF--FGAGNYHLYDEN-----FDLPDYMLSLF-- 410
 DB 222 KHVIYVWDALSNYITLIGDSNDDELNFKWFPADLHLVGRDITRFHTI--YFPIMLMAL 279
 QY 411 -----KKLVGTRKVLMAVQSGSKRKLKLR-----VYLHCTNTDNPYKGGDLTYAINLH 458
 DB 280 DLPLPKQVFGHGWLL--VDGKMSKSGNVADPVVLINFEFTDVRV----- 324
 QY 459 NVTYTLRLPYRFSN-----KQYDKYLRLPLGPHGLSKSVQNLGTLKAVDDOTL 508
 DB 325 -----YLHKEIPFGSDGFENNEIPFKTINSDLANDIG--NLVSRFA--AMIEKYFDGSIQ 375
 QY 509 PPL 511
 DB 376 PPV 378
 RESULT 11
 GYRS CHLPN
 ID GYRS CHLPN STANDARD; PRT; 805 AA.
 AC Q928R3; Q930Q4;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA gyrase subunit B (EC 5.99.1.3).
 GN GYRS OR CPN0275 OR CP0484.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxId=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=CM1029;
 RX MEDLINE=99206606; PubMed=10192338;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., U'terback T., Berry K., Baas S.,
 RA Liner K., Wetzel J., Knouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RA "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RA pneumoniae AR39";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RA from Japan and CWL029 from USA";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
 CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE

INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
DNA RINGS, INCLUDING CATEENANES AND KNOTTED RINGS.
- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
of double-stranded DNA.
- SUBUNIT: MADE UP OF TWO CHAINS: THE A CHAIN IS RESPONSIBLE FOR DNA
BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
ENZYME FORMS AN A2B2 TETRAMER.
- SIMILARITY: Belongs to the type II topoisomerase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL, AE001612; AAD18424.1;
EMBL, AE002210; BAF38314.1;
EMBL, AE002546; BAA98485.1;
PIR, C86525; C86525;
PIR, H72098; H72098;
HSSP, P06982; IAJ6;
PHCI-2DPAGE; Q928R3;
TIGR, CP0484;
InterPro: IPR003594; ATPbind_Arpase.
InterPro: IPR002288; DNA_gyraseb_C.
InterPro: IPR000555; DNA_gyrb.
InterPro: IPR001241; DNA_topoisomII.
InterPro: IPR006171; Toprim_dom.
Pfam: PF00204; DNA_gyraseb_1.
Pfam: PF00986; DNA_gyraseb_C_1.
Pfam: PF02518; HATPase_C_1.
Pfam: PF01751; Toprim_1.
PRINTS: PR00418; TP12FAMILY.
ProDom: PD149633; DNA_gyraseb_C_1.
SMART: SM00433; TOP2C; 1.
TIGRPFAM: TIGR01059; gyrb; 1.
PROSITE: PS00177; TOPOISOMERASE_II; 1.
Topoisomerase; Isomerase; ATP-binding; Complete proteome.
SEQUENCE: 805 AA; 90571 MW; C082DFACCEC71CCC CRC64;
KW

Query Match 3.5%; Score 100.5; DB 1; Length 805;
Best Local Similarity 20.4%; Pred. No. 12;
Matches 121; Conservative 71; Mismatches 169; Indels 231; Gaps 32;

48 PHLVSPS-----FLSVITDANLATDPRFLILGSPKRLTARGSPAVL 92
161 PLQYVSVDROGTEIVFPDPKIFSTCTD-----RSILMKRLRELAFVLRGIT----- 209
93 RRGCTGTDLPIPPKKESTFEERSYQSOVONODICKGSPPYVERKALKEMPYQQLL 152
210 -----IVFDDDDVDVSFDKVTFFYE-----GGQSFVS----- 236
153 REHYOKKPKNSTSRSSVDLYTPANCOSGLDLIFGALNLTADLQMNSSNAQLLDYCS 212
237 ---YLNQKESLDS---EPRIYCGTRVGD-----GEIRFPAALQMSNGSYSLVSYAN 284
213 SKGYNI-SWELGEPNSFLKKADIFINGSQLGSDYIQLHKLRLKSTFFKAK-LYSPDVG 269
285 ---NIPTRCGGTHLNGFSTALTRVIN---TYKKAHILA-----KNNKLLATGEDIL 328
270 QPRKRTAKLAKSLKAGVYIDVTMHVYLNQRTNTRBEPFANPDV-----LDI 318
329 -----REGLTAVISVKNPQFEGQ--TKQKLGSDVSVAAQVQVGEALTI 372
319 PISL-----VQKVF-----QVVESTP-PGK-----K 338
373 FFEENQIARMTDYK/FVAQAPEAKKARELRLRSALDSALPKLLIDCLEKDEKKE 432
339 VMLGETSSAVGGA-----PLSDTPAAGFMWLDKLGSLNKGIEVVMQVFP 386

433 MYVEGSA-GSKAQQRDRFOALIFIRKILINVEKARIQKIFONEIGTIALGCGI 491
387 GAGVYHLDENFDLPYMWLSLPKLVGTGVLM--ASVQSKRRKRV---YLACTN-- 439
492 GADFNL-----SKLRKRI-----IIMTDVDSGSHRTLLLPFYRHMTALI 535
440 -----TDNRYEGDLTLVAINLHVNTKYLRLPYFPNKKQVDKYLRLPGRH--GLLS 490
536 ENECVYIAQPLVY-----VSKKQDRYILSEKEMDSYILM-LGTNSSLIF 581
491 KSV--QANGTLK-----MDDDTLPPLMKRPLRPS-----SLGLPAF 527
582 KSTRERIRGEALSFINVILDVESFINLTKKALPFSEPLEMVEKGIGVPLY 633

RESULT 12
XRNZ_CLOTM STANDARD, PRT, 837 AA.
ID XRNZ_CLOTM
AC P10478
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bndc-1,4-beta-xylanase Z precursor (EC 3.2.1.6) (Xylanase Z)
GN (1,4-beta-D-xylan xylanohydrolase Z).
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OK NCBI_TaxID=1515;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=89008072; PubMed=3139632;
RA Greifet O., Chebrou M., Beguin P.
RT "Nucleotide sequence and deletion analysis of the xylanase gene
(xynZ) of Clostridium thermocellum."
RT J. Bacteriol. 170:4582-4588 (1988).
RN (2)
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS) OF 515-837.
RC STRAIN=NCIB 10682;
RX MEDLINE=95393242; PubMed=7664125;
RA Dominguez R., Soucchon H., Spinelli S., Dauter Z., Wilson K.S.,
RA Chauvaud S., Beguin P., Alzari P.M.
RT "A common protein fold and similar active site in two distinct
families of beta-glycanases."
RT Nat. Struct. Biol. 2:569-576 (1995).
RL
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylanosidic
linkages in xylans.
CC -1- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
CC -1- SIMILARITY: Contains 1 xynZ-type cellulose-binding (CBD) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL, M22624; AAA3286.1;
PIR, A1842; A31842.
PDB, 1XZ; 29-JAN-96.
PDB, 1UTJ; 31-OCT-01.
DR InterPro: IPR00584; CBD_IV.
DR InterPro: IPR005084; CBD_6.
DR InterPro: IPR002105; Dockerin_1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR00801; Esterase_put.

Interpro: IPR001000; Glyco hydro 10.
 Interpro: IPR000379; Ser_esterase.
 Pfam: PF03442; CBM_6, 1. Ser_esterase.
 Pfam: PF00404; Dockerin_1, 2.
 Pfam: PF00756; Esterase_1.
 Pfam: PF00331; Glyco hydro 10, 1.
 PRINTS: PR00134; GLHYDRLASE10.
 SMART: SM00606; CBD-IV, 1.
 SMART: SM00633; Glyco 10, 1.
 PROSITE: PS00018; EF_HAND; UNKNOWN_2.
 PROSITE: PS00448; CLOS_CELLULOSE_RPT, 2.
 PROSITE: PS00591; GLYCOSYL_HYDROL_F10, 1.
 Xylan-degradation; Hydrolase; Glycosidase; Repeat; Signal;
 3D-structure.
 SIGNAL 1 28
 CHAIN 29 837
 ACT_SITE 645 645
 ACT_SITE 754 754
 DOMAIN 328 416
 DOMAIN 430 487
 REPEAT 430 453
 REPEAT 464 487
 REPEAT 783 789
 DISULFID 518 524
 HELIX 525 526
 STRAND 528 533
 STRAND 535 539
 STRAND 540 540
 STRAND 543 552
 STRAND 555 558
 STRAND 561 562
 HELIX 564 567
 STRAND 570 570
 STRAND 571 572
 STRAND 573 573
 STRAND 577 588
 STRAND 589 590
 STRAND 592 599
 STRAND 606 609
 HELIX 610 610
 HELIX 615 632
 STRAND 633 636
 STRAND 639 644
 STRAND 647 647
 STRAND 649 650
 STRAND 654 654
 STRAND 658 663
 STRAND 665 666
 STRAND 667 678
 STRAND 680 681
 STRAND 683 688
 HELIX 696 710
 STRAND 711 712
 STRAND 717 720
 STRAND 723 725
 HELIX 730 745
 STRAND 746 747
 STRAND 749 760
 STRAND 761 762
 STRAND 765 785
 STRAND 787 788
 STRAND 789 794
 STRAND 798 798
 STRAND 799 800
 STRAND 804 807
 STRAND 809 810
 STRAND 811 812
 STRAND 817 817
 STRAND 819 820
 STRAND 823 823
 STRAND 825 834
 HELIX 837 837
 SEQUENCE 837 AA: 92262 MW: DD4C29F04D12B6CD CRC64;

Query Match 3.5% Score 99.5; DB:1; Length 837;
 Best Local Similarity 19.3%; Pred. No. 15;
 Matches 63; Conservative 48; Mismatches 107; Indels 109; Gaps 16;
 145 PROEOLINRE-----HYQKFE-----KNSYRSRSVDLYTFANCGLDIFGNALLR 193
 DB 543 PLYNSILQREFSMVYCEMKEKFDALQPRONVDPFSGQGLAFARNMG-----NR 594
 DB 194 TDLQWNSNAQLLD-----YCSKGYNISMEIGNE-----PNSF 229
 QY 595 GHTLLHNNQNPFWLNGWNNRDSLAVMKNHITTYWTHYKGIYVMDVANECDSDGNGL 654
 DB 230 LKADIFINGSLGSDYDYLQHLKLSFKAKLYGP-----DYQPRKRAK 277
 QY 655 -RSSIWRN-VYGDYLDY-----AFRYAREADPALLFYNDYNIEDLPKSNVFN 703
 QY 278 MKSFLKAGGEVYDSVTWHHYLNGRTATREDFLNPDV-----LDIFISVQKQVNES 332
 DB 704 MKS-MKRGVYIDVGQCHFINGMSPEYLASIDONIKRYAEIGYISFTIDIRIPOS 762
 QY 333 TRPG-----KKWLGRTSSAYGGAPLSDTPAAGFMW--LDKLGLSARMGI 377
 DB 763 ENPATAFYQANNYKELMKICLANPN-----CTFV--MWGFTDKY----- 801
 QY 378 EYVNRQVFRGAGNYHLVDENFDLPDY 404
 DB 802 -TWPGTFPGYGNPLIDSNYNPKPAY 827
 RESULT 13
 ID DYHC HUMAN STANDARD; PRT; 4594 AA.
 AC Q14204; Q92814; Q9Y4G5;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Dynein heavy chain, cytosolic (DYHC). (Cytoplasmic dynein heavy chain
 DE 1) (DHCI) (Fragment).
 GN DNCH1 OR DNECL OR KIAA0325.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 OX (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Ohara O., Nagase T., Kikuno R., Yamakawa H., Nomura N.;
 RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE OF 1078-1974 FROM N.A.
 RX MEDLINE=96234671; PubMed=866668;
 RA Vaisberg E.A., Grissom P.M., McIntosh J.R.;
 RT "Mammalian cells express three distinct dynein heavy chains that are
 RT localized to different cytoplasmic organelles";
 RL J. Cell Biol. 133:831-842 (1996).
 RN (3)
 RP SEQUENCE OF 1832-1972 FROM N.A.
 RX MEDLINE=9403467; PubMed=8227145;
 RA Vaisberg E.A., Koonce M.P., McIntosh J.R.;
 RT "Cytoplasmic dynein plays a role in mammalian mitotic spindle
 RT formation";
 RL J. Cell Biol. 123:849-858 (1993).
 RN (4)
 RP SEQUENCE OF 2508-4594 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohita M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro";
 RL DNA Res. 4:141-150 (1997).

CC -1- FUNCTION: Cytoplasmic dynein acts as a motor for the intracellular
CC retrograde motility of vesicles and organelles along microtubules.
CC Dynein has ATPase activity; the force-producing power stroke is
CC thought to occur on release of ADP.
CC -1- SUBUNIT: Consists of at least two heavy chains and a number of
CC intermediate and light chains.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- DOMAIN: Dynein heavy chains probably consist of an N-terminal stem
CC (which binds cargo and interacts with other dynein components),
CC and the head or motor domain. The motor contains six tandemly-
CC linked AAA domains in the head, which form a ring. A stalk-like
CC structure (formed by two of the coiled coil domains) protrudes
CC between AAA 4 and AAA 5 and terminates in a microtubule-binding
CC site. A seventh domain may also contribute to this ring; it is not
CC clear whether the N-terminus or the C-terminus forms this extra
CC domain. There are four well-conserved and two non-conserved ATPase
CC sites, one per AAA domain. Probably only one of these (within AAA
CC 1) actually hydrolyzes ATP, the others may serve a regulatory
CC function.
CC -1- SIMILARITY: Belongs to the dynein heavy chain family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB002323; BAA20783.2;
CC EMBL; U53530; AAB09727.1;
CC EMBL; L23958; AAB16065.1;
CC PIR; A49019; A49019;
CC PIR; G02529; G02529;
CC GeneW; HGNC:2961; DNCH1.
CC MIM; 600112;
CC GO; GO:0005868; C:cytoplasmic dynein complex; NAS.
CC GO; GO:0008577; F:dynein ATPase activity; NAS.
CC GO; GO:000377; F:microtubule motor activity; NAS.
CC GO; GO:0007052; P:mitotic spindle assembly; NAS.
CC InterPro; IPR005613; AIP3.
CC InterPro; IPR004273; Dynein_heavy.
CC Pfam; PF03028; Dynein_heavy; 1.
CC Pfam; PF03915; AIP3; 1.
CC SMART; SM00382; AAA; 4.
CC Motor protein; Dynein; Microtubules; ATP-binding; Repeat; Coiled coil.
CC NON_TER 1 1815 STEM (BY SIMILARITY).
CC DOMAIN 1 1815 AAA 1 (BY SIMILARITY).
CC DOMAIN 1816 2047 AAA 2 (BY SIMILARITY).
CC DOMAIN 2128 2400 AAA 3 (BY SIMILARITY).
CC DOMAIN 2504 2753 AAA 4 (BY SIMILARITY).
CC DOMAIN 2847 3116 STALK (BY SIMILARITY).
CC DOMAIN 3137 3448 AAA 5 (BY SIMILARITY).
CC DOMAIN 3501 3730 AAA 6 (BY SIMILARITY).
CC DOMAIN 3953 4169 COILED COIL (POTENTIAL).
CC DOMAIN 129 150 COILED COIL (POTENTIAL).
CC DOMAIN 403 426 COILED COIL (POTENTIAL).
CC DOMAIN 491 514 COILED COIL (POTENTIAL).
CC DOMAIN 1119 1200 COILED COIL (POTENTIAL).
CC DOMAIN 1305 1321 COILED COIL (POTENTIAL).
CC DOMAIN 3137 3223 COILED COIL (POTENTIAL).
CC DOMAIN 3344 3448 COILED COIL (POTENTIAL).
CC DOMAIN 3685 3748 COILED COIL (POTENTIAL).
CC NP_BIND 1854 1861 ATP (POTENTIAL).
CC NP_BIND 2172 2179 ATP (POTENTIAL).
CC NP_BIND 2543 2550 ATP (POTENTIAL).
CC NP_BIND 2885 2892 ATP (POTENTIAL).
CC NP_BIND 1726 1727 LH -> SD (IN REF. 2).
CC CONFLICT 1889 1889 M -> R (IN REF. 2).
CC CONFLICT 1973 1973 R -> N (IN REF. 2).
CC CONFLICT 1977 1977 H -> Q (IN REF. 4).
CC SEQUENCE 4594 AA; 527282 MW; B6F8DA7F717C169 CRC64;

Query Match 3.5%; Score 99; DB 1; Length 4594;
Best Local Similarity 19.7%; Pred. No. 2e+02;
Matches 113; Conservative 82; Mismatches 172; Indels 208; Gaps 27;
QY 41 IDFFQBEFLHVSFSLVTDANLADPRLILGSKLTARGLSPATVREGCTTD 100
DB 1231 IDQMGQVSVOPKRNQNDALNQLKSF -PARLQVASYEPQRLNGYMKIN 1345
QY 101 FLIPDKESFEESVYQSVNQ -DICKXGISPPDV - 136
DB 1336 MIVIEFLKSEA-LKOR-HKQMLKRLAHNVVSELTGLQIMVYDQNKNAIKVDVLVAQG 1403
QY 137 ---EE---KRLKLEPYQOULLREHYQKPKNSTYSSVDVLYTFANCSGLDIFGLN 189
DB 1404 EWALEEFKQJREVNTWTELDLV -NYQNKCR -LIRGMD 1439
QY 130 ALKRPADLQNNSSNAQILLDYCSSKGVNI -SWELENEPSFLKADIFIN - 238
DB 1440 DLFNPKKHINSVSAMKLSPP -YKVFEDALSWB -DKLRIMALFPVWIDVQRRW 1492
QY 239 ---GS---QLGEDYIQLKLRKSTFKNAKLPGDVGOPRRKT 275
DB 1493 VYLSGIFPGSADIKHLPEVETQROSIETFLALMKKYSKPLVMDVNLIGVQSLRL 1552
QY 276 AKMLKSPFKAGGEVY -DSVTWHYYLNGRTATREDPLNPDVDF -ISSYQVVFQ 328
DB 1553 ADLLKQIKQALGEYLERSSFPFRYFVG -DEDLLEIIGNSKNVAKLQKH - 1602
QY 329 VVESTRPGRKVMLETSAYGGAPLSDTFAAGFMWLDKLSRMKIEVVMRVVFQGA 388
DB 1603 ---KKMFAGVSSI -ILNEDNSV -VGISSEEEVWFKPP - 1637
QY 389 GNVHLVDENFPLDPYMLSLFKLVGTGVLMASVQSKRRRLRYVLTCTNDNRYREG 448
DB 1638 ---VSTTEPKINEMWLTVEKEN -RVTLAKLLASVTEVLEFGATSIDPRTY - 1686
QY 449 DLTVALNMLHVTYKTLRLPYEPNSKQVDKY - - - - -LRLPGP - 485
DB 1687 ---IT - - - - -WIDKYOQLVVLVAQAIAVSHVETALSSMGCGDA 1723
QY 486 ---HGLLSK-SYQNLGLTLKAVDDOTLPLMEKPL 516
DB 1724 APLHSLVLSNVETLNLVLDVSLVMEQ -PPLRRRLK 1756
RESULT 14
DHC RAT STANDARD; PRT; 4644 AA.
AC P38650; Q63178; 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dynein heavy chain, cytosolic (DHC) (Cytoplasmic dynein heavy chain)
DE (Map 1C).
GN DNCH1 OR DNCH1 OR DNEC1 OR MAP1C.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Missar; TISSUE=Brain;
RX MEDLINE=93376715; PubMed=7690137;
RA Zhang Z., Tanaka Y., Nonaka S., Aizawa H., Kawasaki H., Nakata T.,
RT "The primary structure of rat brain (cytoplasmic) dynein heavy chain,
RT a cytoplasmic motor enzyme.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:7928-7932(1993).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=93264075; PubMed=7684232;
RA Mikami A., Paschal B.M., Mazumdar M., Vallée R.B.;

FT	CONFLICT	4366	4366	F -> S (IN REF. 2).
FT	CONFLICT	4511	4511	A -> G (IN REF. 2).
SO	SEQUENCE	4644	AA: 532240 MW: 8C6ABDBEDF875D82	CRC64;
Query Match 3.5%; Score 99; DB 1; Length 4644;				
Best Local Similarity 19.5%; Pred. No. 2e+02;				
Matches 109; Conservative 85; Mismatches 188; Indels 178; Gaps 25;				
QY	41	LDPEPTDRLHVPSPFLSVITIDANLATDPRFLILGSPKRLTLAGSPAYLRFQGTKT	100	
DB	1341	IDQKDEQPMWVQVRKRLKQMLDGLNLQKNF----	PARLRQYASVEFYQRLKLGWKIN	1395
QY	101	FLIDPRKKESTFEERSYWCQVNO-----	DICKGSIIPDY-----	136
DB	1396	MLVIELKSEA-LKKR-HMKQIMKRLHVMVWSBLTQIQWDVDLQKNEATYKVDLLVAOG	1453	
QY	137	-----EE-----KLRLMPYQEQDLLREHQKFKNSTGRSSVDVLYTFANSGDILIGLN	189	
DB	1454	EMALIEFLKQIREYMTYBIDLV-NKQKCR-----	LINGWD	1489
QY	190	ALATADLQNMSSNAQDLIDYCCKGYNI-----	SMELGNBPSFLKADIFIN-----	238
DB	1490	DLFNKVKXEHINSVSAKLSPY-----	YKVEBDALSP--DKLNRIALFDWVITDQGRW	1542
QY	239	-----GS-----	QLGDDYIQLHLKLRSTFKNAQLYGPDVQGPRT	275
DB	1543	VYLEGIFPTGGADIKHLPLVETORFQSIETEPFLMKVKYSPLVMDVLTNIQVGRSLERL	1602	
QY	276	AKMLKSLFKKGGVYI--DSVTWHNYLNGTATREDPLNDVLDIF-----	ISSGVQVFG	328
DB	1603	ADLGLKTIQKALGVELERERSFPREFYVG-----	DEDLLEITIGNSKNVAKLQHF-	1652
QY	329	VESTREPKTVLGETISSAYGAGAPLISDTPAAGFMWLDLGLSARMGIEVMKQVFEGA	388	
DB	1653	-----KMPFAGVSIIINEDSSV-----	LGISSRGEEMFTFP----	1687
QY	389	GNTHLVDENDPRLPDYVLSLFLPKLVGTQVLMASVQSKRKRLRVYLHCTNTDNPY--	445	
DB	1688	-----VSITEHPKINEMWLTIVKEM--RVTLAKTLAESYVEVEIKPKATSIDENTYITW	1739	
QY	446	-----KEGDLTYAINL---NHVTKYLRPLPYFSNKQVNDKYLLRPLGPHGLSKSYQNLGT	499	
DB	1740	IDKYQAQLVYLSAQIAIASEVENALSNVGGGN-----	VGPLQSVLSNVE--VT	1786
QY	500	LKAWVDQTL--PPLMEKPL	516	
DB	1787	LNTVLABSVLMEQEPPLRRKL	1806	
RESULT 15				
T3MH_HABIN				
ID	T3MH_HABIN	STANDARD;	PRT;	629 AA.
AC	P7136;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Puative type III restriction-modification system HindVIP enzyme mod			
DE	(EC 2.1.1.72) (HindVIP-methyltransferase) (M.HindVIP).			
GN	H11056.			
OC	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
OX	NCBI_TaxId=727;			
RE	SEQUENCE FROM N.A.			
RC	STRAIN=Rd / KW20 / ATCC 51907;			
RX	MEDLINE=95350630; PubMed=7542800;			
RA	Pfeleishman R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,			
RA	Weidman J.F., Phillips C.A., Spriggs T., Heddiom E., Cotton M.D.,			
RA	Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,			

RA Fine, L.D., Fritschman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm, C.L., McDonald, L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.,
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -1- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD. MOD IS
CC N HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: WITH OTHER TYPE III MOD PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32786; AAC22721.1; -
CC PIR: C64180; C64180
DR REBASE: 3701; M.HINDORF1056P.
DR TIGR: H11056;
DR InterPro: IPR001091; CNA_Mettransf.
DR InterPro: IPR002295; D2IN6_Mtase.
DR InterPro: IPR002941; N6/N4_Mtase.
DR InterPro: IPR002052; N6_Mtase.
DR Pfam: PF01555; N6_N4_Mtase; 1.
DR PRINTS: PR00506; D2IN6MTFRASE.
DR PRINTS: PR00508; S2IN4MTFRASE.
DR PROSITE: PS00092; N6_MTASE; 1.
KW Hypothetical protein; Transferrase; Methyltransferase; DNA-binding;
KW Restriction system; Complete proteome.
SQ SEQUENCE 629 AA; 71845 MW; 93ADAD909DB41B84 CRC64;

Query Match: 3.5%; Score 98.5; DB 1; Length 629;
Best Local Similarity 21.8%; Pred. No. 12;
Matches 67; Conservative 56; Mismatches 107; Indels 77; Gaps 19;

QY 99 TDFL-IPDKKSTFEERYSQVNOVODICKYGIIPDVEEKLRLWPQ-----EQL 150
DB 242 TEFIVFSKRRRLPLNPLYOKTEVNEFIKNY-----EDSGKSWKYQVLIIDLGSKI 293
QY 151 LIREHYOKKFKSTYSRSSVDVLYTFANCSGL--DLIFG--LNALLRTADLQ----- 198
DB 294 LHEE--KGFKYTHPNAMQTSIVFSDQDNLKSKILLIYTSKHYRTTMAOSSIRSKII 351
QY 199 ---WNSSNAQLLDYCSSKQYN-----ISWELGNEPNSFLKADIFING----- 239
DB 352 EDLYSIKNGIVSIEYIPQKKNAGNLIIEFYVANSNK--DMFMFLSDMLIKKKKYFYLOKV 410
QY 240 SOLGSDYIQHLGLKKS---TFKNAKLYGPDVGGPRRTAKMLKSLKAGSGVIDSVTWH 296
DB 411 NTLWMD-IQYNNLNKEGGYIDFKNGK-----KPEALLRIIDMTTKBGDIVLD----- 457
QY 297 HYLL-NGRTATREDFANPDVLDI---FTS--SVOKVFOVESTSPG--KKY-MLGSTSS 346
DB 458 -YHLSSGTTAAVAHAKMROYIGIEOMDYIETTLAVERLKKVVIDGEOGGLSKAVANOGGGBF 516
QY 347 AVGGAP 353
DB 517 VYABLAP 523

Search completed: October 22, 2003; 20:24:56
Job time: 29 secs